

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C07K 14/415, C12N 15/29, C12Q 1/68</b>		<b>A1</b>	(11) International Publication Number: <b>WO 95/28423</b>
			(43) International Publication Date: 26 October 1995 (26.10.95)
(21) International Application Number: <b>PCT/US95/04589</b>		(72) Inventors: AUSUBEL, Frederick, M.; 271 Lake Avenue, Newton, MA 02161 (US). STASKAWICZ, Brian, J.; 18945 Marciel Court, Castro Valley, CA 94546 (US). BENT, Andrew, F.; 308 West Iowa Street, Urbana, IL 61801 (US). DAHLBECK, Douglas; 18851 Lenross Court, Castro Valley, CA 94546 (US). KATAGIRI, Fumiaki; Seven Hardan Road, Somerville, MA 02144 (US). KUNKEL, Barbara, N.; 6042 Pershing Avenue, St. Louis, MO 63112 (US). MINDRINOS, Michael, N.; 400 Brookline Avenue, Boston, MA 02215 (US). YU, Guo-Liang; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). BAKER, Barbara; 804 Vista Heights Road, Richmond, CA 94805 (US). ELLIS, Jeffrey; 10 Gibbs Place, Weetangera 2614 (AU). SALMERON, John; 1308 Blackberry Lane, Hillsborough, NC 27278 (US).	
(22) International Filing Date: 13 April 1995 (13.04.95)		(74) Agent: CLARK, Paul, T.; Fish & Richardson, 225 Franklin Street, Boston, MA 02110-2804 (US).	
(30) Priority Data: 08/227,360 13 April 1994 (13.04.94) US 08/310,912 22 September 1994 (22.09.94) US		(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).	
(71) Applicants: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02114 (US). THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612-3550 (US). THE UNITED STATES OF AMERICA, represented by THE SECRETARY OF AGRICULTURE [US/US]; United States Department of Agriculture, 14th & Independence Avenue, SW, Washington, DC 20250 (US). COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION [AU/AU]; 407 Royal Parade, Parkville, VIC 3052 (AU).		Published With international search report.	
(54) Title: <b>RPS GENE FAMILY, PRIMERS, PROBES, AND DETECTION METHODS</b>			
(57) Abstract  Disclosed is substantially pure DNA encoding an <i>Arabidopsis thaliana</i> Rps2 polypeptide; substantially pure Rps2 polypeptide; and methods of using such DNA to express the Rps2 polypeptide in plant cells and whole plants to provide, in transgenic plants, disease resistance to pathogens. Also disclosed are conserved regions characteristic of the RPS family and primers and probes for the identification and isolation of additional RPS disease-resistance genes.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

- 1 -

RPS GENE FAMILY, PRIMERS, PROBES, AND DETECTION METHODSStatement as to Federally Sponsored Research

5           This invention was made in part with Government funding and the Government therefore has certain rights in the invention.

Background of the Invention

10           The invention relates to recombinant plant nucleic acids and polypeptides and uses thereof to confer disease resistance to pathogens in transgenic plants.

          Plants employ a variety of defensive strategies to combat pathogens. One defense response, the so-called hypersensitive response (HR), involves rapid localized  
15       necrosis of infected tissue. In several host-pathogen interactions, genetic analysis has revealed a gene-for-gene correspondence between a particular avirulence (avr) gene in an avirulent pathogen that elicits an HR in a host possessing a particular resistance gene.

20

Summary of the Invention

          In general, the invention features substantially pure DNA (for example, genomic DNA, cDNA, or synthetic DNA) encoding an Rps polypeptide as defined below. In related aspects, the invention also features a vector, a  
25       cell (e.g., a plant cell), and a transgenic plant or seed thereof which includes such a substantially pure DNA encoding an Rps polypeptide.

          In preferred embodiments, an RPS gene is the RPS2 gene of a plant of the genus *Arabidopsis*. In various  
30       preferred embodiments, the cell is a transformed plant cell derived from a cell of a transgenic plant. In related aspects, the invention features a transgenic plant containing a transgene which encodes an Rps

- 2 -

polypeptide that is expressed in plant tissue susceptible to infection by pathogens expressing the *avrRpt2* avirulence gene or pathogens expressing an avirulence signal similarly recognized by an *Rps* polypeptide.

5 In a second aspect, the invention features a substantially pure DNA which includes a promoter capable of expressing the *RPS2* gene in plant tissue susceptible to infection by bacterial pathogens expressing the *avrRpt2* avirulence gene.

10 In preferred embodiments, the promoter is the promoter native to an *RPS* gene. Additionally, transcriptional and translational regulatory regions are preferably native to an *RPS* gene.

The transgenic plants of the invention are  
15 preferably plants which are susceptible to infection by a pathogen expressing an avirulence gene, preferably the *avrRpt2* avirulence gene. In preferred embodiments the transgenic plant is from the group of plants consisting of but not limited to *Arabidopsis*, tomato, soybean, bean,  
20 maize, wheat and rice.

In another aspect, the invention features a method of providing resistance in a plant to a pathogen which involves: (a) producing a transgenic plant cell having a transgene encoding an *Rps2* polypeptide wherein the  
25 transgene is integrated into the genome of the transgenic plant and is positioned for expression in the plant cell; and (b) growing a transgenic plant from the transgenic plant cell wherein the *RPS2* transgene is expressed in the transgenic plant.

30 In another aspect, the invention features a method of detecting a resistance gene in a plant cell involving: (a) contacting the *RPS2* gene or a portion thereof greater than 9 nucleic acids, preferably greater than 18 nucleic acids in length with a preparation of genomic DNA from  
35 the plant cell under hybridization conditions providing

- 3 -

detection of DNA sequences having about 50% or greater sequence identity to the DNA sequence of Fig. 2 encoding the Rps2 polypeptide.

In another aspect, the invention features a method  
5 of producing an Rps2 polypeptide which involves: (a) providing a cell transformed with DNA encoding an Rps2 polypeptide positioned for expression in the cell; (b) culturing the transformed cell under conditions for expressing the DNA; and (c) isolating the Rps2  
10 polypeptide.

In another aspect, the invention features substantially pure Rps2 polypeptide. Preferably, the polypeptide includes a greater than 50 amino acid sequence substantially identical to a greater than 50  
15 amino acid sequence shown in Fig. 2, open reading frame "a". Most preferably, the polypeptide is the *Arabidopsis thaliana* Rps2 polypeptide.

In another aspect, the invention features a method of providing resistance in a transgenic plant to  
20 infection by pathogens which do not carry the avrRpt2 avirulence gene wherein the method includes: (a) producing a transgenic plant cell having transgenes encoding an Rps2 polypeptide as well as a transgene encoding the avrRpt2 gene product wherein the transgenes  
25 are integrated into the genome of the transgenic plant; are positioned for expression in the plant cell; and the avrRpt2 transgene and, if desired, the RPS2 gene, are under the control of regulatory sequences suitable for controlled expression of the gene(s); and (b) growing a  
30 transgenic plant from the transgenic plant cell wherein the RPS2 and avrRpt2 transgenes are expressed in the transgenic plant.

In another aspect, the invention features a method of providing resistance in a transgenic plant to  
35 infection by pathogens in the absence of avirulence gene

- 4 -

expression in the pathogen wherein the method involves:

(a) producing a transgenic plant cell having integrated in the genome a transgene containing the *RPS2* gene under the control of a promoter providing constitutive

- 5 expression of the *RPS2* gene; and (b) growing a transgenic plant from the transgenic plant cell wherein the *RPS2* transgene is expressed constitutively in the transgenic plant.

In another aspect, the invention features a method  
10 of providing controllable resistance in a transgenic plant to infection by pathogens in the absence of avirulence gene expression in the pathogen wherein the method involves: (a) producing a transgenic plant cell having integrated in the genome a transgene containing  
15 the *RPS2* gene under the control of a promoter providing controllable expression of the *RPS2* gene; and (b) growing a transgenic plant from the transgenic plant cell wherein the *RPS2* transgene is controllably expressed in the transgenic plant. In preferred embodiments, the *RPS2*  
20 gene is expressed using a tissue-specific or cell type-specific promoter, or by a promoter that is activated by the introduction of an external signal or agent, such as a chemical signal or agent.

In other aspects, the invention features a  
25 substantially pure oligonucleotide including one or a combination of the sequences:

5' GGNATGGGNGGNNTNGGNAARACNAC 3', [SEQ ID NO: 158]  
wherein N is A, T, G, or C; and R is A or G;

5' NARNGGNARNCC 3', [SEQ ID NO: 169] wherein N is  
30 A, T, G or C; and R is A or G;

5'NCGNGWNGTNAKDAWNCNGNA 3', [SEQ ID NO: 159]  
wherein N is A, T, G or C; W is A or T; D is A, G, or T;  
and K is G or T;

- 5 -

5' GGWNTBGGWAARACHAC 3', [SEQ ID NO: 160] wherein N is A, T, G or C; R is G or A; B is C, G, or T; H is A, C, or T; and W is A or T;

5' TYGAYGAYRTBKRBRA 3', [SEQ ID NO: 163] wherein R is G or A; B is C, G, or T; D is A, G, or T; Y is T or C; and K is G or T;

5' TYCCAVAYRTCRTCNA 3', [SEQ ID NO: 164] wherein N is A, T, G or C; R is G or A; V is G or C or A; and Y is T or C;

10 5' GGWYTBCCWYTBGCHYT 3', [SEQ ID NO: 170] wherein B is C, G, or T; H is A, C, or T; W is A or T; and Y is T or C;

5' ARDGCVARWGGVARNCC 3', [SEQ ID NO: 171] wherein N is A, T, G or C; R is G or A; W is A or T; D is A, G, or T; and V is G, C, or A; and

5' ARRTTRTCRTADSWRAWYTT 3', [SEQ ID NO: 174] wherein R is G or A; W is A or T; D is A, G, or T; S is G or C; and Y is C or T.

In other aspects, the invention features a recombinant plant gene including one or a combination of the DNA sequences:

5' GGNATGGGNGGNNTNGGNAARACNAC 3', [SEQ ID NO: 162] wherein N is A, T, G or C; and R is A or G;

5' NARNGGNARNCC 3', [SEQ ID NO: 169] wherein N is A, T, G or C; and R is A or G;

5' NCGNGWNGTNAKDAWNCNGA 3', [SEQ ID NO: 167] wherein N is A, T, G or C; W is A or T; D is A, G or T; and K is G or T.

In another aspect, the invention features a substantially pure plant polypeptide including one or a combination of the amino acid sequences:

Gly Xaa<sub>1</sub> Xaa<sub>2</sub> Gly Xaa<sub>3</sub> Gly Lys Thr Thr Xaa<sub>4</sub> Xaa<sub>5</sub>, [SEQ ID NO: 191] wherein Xaa<sub>1</sub> is Met or Pro; Xaa<sub>2</sub> is Gly or Pro; Xaa<sub>3</sub> is Ile, Leu, or Val; Xaa<sub>4</sub> is Ile, Leu, or Thr; and Xaa<sub>5</sub> is Ala or Met;

- 6 -

Xaa<sub>1</sub> Xaa<sub>2</sub> Xaa<sub>3</sub> Leu Xaa<sub>4</sub> Xaa<sub>5</sub> Xaa<sub>6</sub> Asp Asp Xaa<sub>7</sub>  
Xaa<sub>8</sub>, [SEQ ID NO: 192]

wherein Xaa<sub>1</sub> is Phe or Lys; Xaa<sub>2</sub> is Arg or Lys; Xaa<sub>3</sub> is  
Ile, Val, or Phe; Xaa<sub>4</sub> is Ile, Leu, or Val; Xaa<sub>5</sub> is Ile or  
5 Leu; Xaa<sub>6</sub> is Ile or Val; Xaa<sub>7</sub> is Ile, Leu, or Val; and  
Xaa<sub>8</sub> is Asp or Trp;

Xaa<sub>1</sub> Xaa<sub>2</sub> Xaa<sub>3</sub> Xaa<sub>4</sub> Xaa<sub>5</sub> Thr Xaa<sub>6</sub> Arg, [SEQ ID NO:  
193]

wherein Xaa<sub>1</sub> is Ser or Cys; Xaa<sub>2</sub> is Arg or Lys; Xaa<sub>3</sub> is  
10 Phe, Ile, or Val; Xaa<sub>4</sub> is Ile, or Met; Xaa<sub>5</sub> is Ile, Leu,  
or Phe; Xaa<sub>6</sub> is Ser, Cys, or Thr;

Gly Leu Pro Leu Xaa<sub>1</sub> Xaa<sub>2</sub> Xaa<sub>3</sub> Xaa<sub>4</sub>, [SEQ ID NO.:  
194]

wherein Xaa<sub>1</sub> is Thr, Ala, or Ser; Xaa<sub>2</sub> is Leu or Val; Xaa<sub>3</sub>  
15 is Ile, Val, or Lys; and Xaa<sub>4</sub> is Val or Thr; and

Xaa<sub>1</sub> Xaa<sub>2</sub> Ser Tyr Xaa<sub>3</sub> Xaa<sub>4</sub> Leu, [SEQ ID NO: 195]  
wherein Xaa<sub>1</sub> is Lys or Gly; Xaa<sub>2</sub> is Ile or Phe; Xaa<sub>3</sub> is  
Asp or Lys; and Xaa<sub>4</sub> is Ala, Gly, or Asn.

In another aspect, the invention features a method  
20 of isolating a disease-resistance gene or fragment  
thereof from a plant cell, involving: (a) providing a  
sample of plant cell DNA; (b) providing a pair of  
oligonucleotides having sequence homology to a conserved  
region of an RPS disease-resistance gene; (c) combining  
25 the pair of oligonucleotides with the plant cell DNA  
sample under conditions suitable for polymerase chain  
reaction-mediated DNA amplification; and (d) isolating  
the amplified disease-resistance gene or fragment  
thereof.



- 7 -

In preferred embodiments, the amplification is carried out using a reverse-transcription polymerase chain reaction, for example, the RACE method

In another aspect, the invention features a method  
5 of identifying a plant disease-resistance gene in a plant cell, involving: (a) providing a preparation of plant cell DNA (for example, from the plant genome); (b) providing a detectably-labelled DNA sequence (for example, prepared by the methods of the invention) having  
10 homology to a conserved region of an RPS gene; (c) contacting the preparation of plant cell DNA with the detectably-labelled DNA sequence under hybridization conditions providing detection of genes having 50% or greater sequence identity; and (d) identifying a disease-  
15 resistance gene by its association with the detectable label.

In another aspect, the invention features a method of isolating a disease-resistance gene from a recombinant plant cell library, involving: (a) providing a  
20 recombinant plant cell library; (b) contacting the recombinant plant cell library with a detectably-labelled gene fragment produced according to the PCR method of the invention under hybridization conditions providing detection of genes having 50% or greater sequence  
25 identity; and (c) isolating a member of a disease-resistance gene by its association with the detectable label.

In another aspect, the invention features a method of isolating a disease-resistance gene from a recombinant  
30 plant cell library, involving: (a) providing a recombinant plant cell library; (b) contacting the recombinant plant cell library with a detectably-labelled RPS oligonucleotide of the invention under hybridization conditions providing detection of genes having 50% or  
35 greater sequence identity; and (c) isolating a disease-

- 8 -

resistance gene by its association with the detectable label.

In another aspect, the invention features a recombinant plant polypeptide capable of conferring disease-resistance wherein the plant polypeptide includes a P-loop domain or nucleotide binding site domain. Preferably, the polypeptide further includes a leucine-rich repeating domain.

In another aspect, the invention features a recombinant plant polypeptide capable of conferring disease-resistance wherein the plant polypeptide contains a leucine-rich repeating domain.

In another aspect, the invention features a plant disease-resistance gene isolated according to the method involving: (a) providing a sample of plant cell DNA; (b) providing a pair of oligonucleotides having sequence homology to a conserved region of an RPS disease-resistance gene; (c) combining the pair of oligonucleotides with the plant cell DNA sample under conditions suitable for polymerase chain reaction-mediated DNA amplification; and (d) isolating the amplified disease-resistance gene or fragment thereof.

In another aspect, the invention features a plant disease-resistance gene isolated according to the method involving: (a) providing a preparation of plant cell DNA; (b) providing a detectably-labelled DNA sequence having homology to a conserved region of an RPS gene; (c) contacting the preparation of plant cell DNA with the detectably-labelled DNA sequence under hybridization conditions providing detection of genes having 50% or greater sequence identity; and (d) identifying a disease-resistance gene by its association with the detectable label.

In another aspect, the invention features a plant disease-resistance gene according to the method

- 9 -

involving: (a) providing a recombinant plant cell library; (b) contacting the recombinant plant cell library with a detectably-labelled RPS gene fragment produced according to the method of the invention under  
5 hybridization conditions providing detection of genes having 50% or greater sequence identity; and (c) isolating a disease-resistance gene by its association with the detectable label.

In another aspect, the invention features a method  
10 of identifying a plant disease-resistance gene involving: (a) providing a plant tissue sample; (b) introducing by biolistic transformation into the plant tissue sample a candidate plant disease-resistance gene; (c) expressing the candidate plant disease-resistance gene within the  
15 plant tissue sample; and (d) determining whether the plant tissue sample exhibits a disease-resistance response, whereby a response identifies a plant disease-resistance gene.

Preferably, the plant tissue sample is either  
20 leaf, root, flower, fruit, or stem tissue; the candidate plant disease-resistance gene is obtained from a cDNA expression library; and the disease-resistance response is the hypersensitive response.

In another aspect, the invention features a plant  
25 disease-resistance gene isolated according to the method involving: (a) providing a plant tissue sample; (b) introducing by biolistic transformation into the plant tissue sample a candidate plant disease-resistance gene; (c) expressing the candidate plant disease-resistance  
30 gene within the plant tissue sample; and (d) determining whether the plant tissue sample exhibits a disease-resistance response, whereby a response identifies a plant disease-resistance gene.

In another aspect, the invention features a  
35 purified antibody which binds specifically to an rps

- 10 -

family protein. Such an antibody may be used in any standard immunodetection method for the identification of an RPS polypeptide.

In another aspect, the invention features a DNA  
5 sequence substantially identical to the DNA sequence shown in Figure 12.

In another aspect, the invention features a substantially pure polypeptide having a sequence substantially identical to a Prf amino acid sequence  
10 shown in Figure 5 (A or B).

By "disease resistance gene" is meant a gene encoding a polypeptide capable of triggering the plant defense response in a plant cell or plant tissue. An RPS gene is a disease resistance gene having about 50% or  
15 greater sequence identity to the RPS2 sequence of Fig. 2 or a portion thereof. The gene, RPS2, is a disease resistance gene encoding the Rps2 disease resistance polypeptide from *Arabidopsis thaliana*.

By "polypeptide" is meant any chain of amino  
20 acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation).

By "substantially identical" is meant a polypeptide or nucleic acid exhibiting at least 50%, preferably 85%, more preferably 90%, and most preferably  
25 95% homology to a reference amino acid or nucleic acid sequence. For polypeptides, the length of comparison sequences will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably 35 amino acids.  
30 For nucleic acids, the length of comparison sequences will generally be at least 50 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 110 nucleotides.

Sequence identity is typically measured using  
35 sequence analysis software (e.g., Sequence Analysis

- 11 -

Software Package of the Genetics Computer Group,  
University of Wisconsin Biotechnology Center, 1710  
University Avenue, Madison, WI 53705). Such software  
matches similar sequences by assigning degrees of  
5 homology to various substitutions, deletions,  
substitutions, and other modifications. Conservative  
substitutions typically include substitutions within the  
following groups: glycine alanine; valine, isoleucine,  
leucine; aspartic acid, glutamic acid, asparagine,  
10 glutamine; serine, threonine; lysine, arginine; and  
phenylalanine, tyrosine.

By a "substantially pure polypeptide" is meant an  
Rps2 polypeptide which has been separated from components  
which naturally accompany it. Typically, the polypeptide  
15 is substantially pure when it is at least 60%, by weight,  
free from the proteins and naturally-occurring organic  
molecules with which it is naturally associated.  
Preferably, the preparation is at least 75%, more  
preferably at least 90%, and most preferably at least  
20 99%, by weight, Rps2 polypeptide. A substantially pure  
Rps2 polypeptide may be obtained, for example, by  
extraction from a natural source (e.g., a plant cell); by  
expression of a recombinant nucleic acid encoding an Rps2  
polypeptide; or by chemically synthesizing the protein.  
25 Purity can be measured by any appropriate method, e.g.,  
those described in column chromatography, polyacrylamide  
gel electrophoresis, or by HPLC analysis.

A protein is substantially free of naturally  
associated components when it is separated from those  
30 contaminants which accompany it in its natural state.  
Thus, a protein which is chemically synthesized or  
produced in a cellular system different from the cell  
from which it naturally originates will be substantially  
free from its naturally associated components.  
35 Accordingly, substantially pure polypeptides include

- 12 -

those derived from eukaryotic organisms but synthesized in *E. coli* or other prokaryotes.

By "substantially pure DNA" is meant DNA that is free of the genes which, in the naturally-occurring  
5 genome of the organism from which the DNA of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of  
10 a prokaryote or eukaryote; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA which is part of a hybrid gene encoding  
15 additional polypeptide sequence.

By "transformed cell" is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding (as used herein) an Rps2 polypeptide.

20 By "positioned for expression" is meant that the DNA molecule is positioned adjacent to a DNA sequence which directs transcription and translation of the sequence (i.e., facilitates the production of, e.g., an Rps2 polypeptide, a recombinant protein or a RNA  
25 molecule).

By "reporter gene" is meant a gene whose expression may be assayed; such genes include, without limitation,  $\beta$ -glucuronidase (GUS), luciferase, chloramphenicol transacetylase (CAT), and  $\beta$ -  
30 galactosidase.

By "promoter" is meant minimal sequence sufficient to direct transcription. Also included in the invention are those promoter elements which are sufficient to render promoter-dependent gene expression controllable  
35 for cell-type specific, tissue-specific or inducible by

- 13 -

external signals or agents; such elements may be located in the 5' or 3' regions of the native gene.

By "operably linked" is meant that a gene and a regulatory sequence(s) are connected in such a way as to  
5 permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence(s).

By "plant cell" is meant any self-propagating cell bounded by a semi-permeable membrane and containing a  
10 plastid. Such a cell also requires a cell wall if further propagation is desired. Plant cell, as used herein includes, without limitation, algae, cyanobacteria, seeds suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes,  
15 sporophytes, pollen, and microspores.

By "transgene" is meant any piece of DNA which is inserted by artifice into a cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or  
20 entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism.

By "transgenic" is meant any cell which includes a DNA sequence which is inserted by artifice into a cell  
25 and becomes part of the genome of the organism which develops from that cell. As used herein, the transgenic organisms are generally transgenic plants and the DNA (transgene) is inserted by artifice into the nuclear or plastidic genome.

30 By "pathogen" is meant an organism whose infection into the cells of viable plant tissue elicits a disease response in the plant tissue.

By an "RPS disease-resistance gene" is meant any member of the family of plant genes characterized by  
35 their ability to trigger a plant defense response and

- 14 -

having at least 20%, preferably 30%, and most preferably 50% amino acid sequence identity to one of the conserved regions of one of the RPS members described herein (i.e., either the RPS2, L6, N, or Prf genes). Representative members of the RPS gene family include, without limitation, the rps2 gene of *Arabidopsis*, the L6 gene of flax, the Prf gene of tomato, and the N gene of tobacco.

By "conserved region" is meant any stretch of six or more contiguous amino acids exhibiting at least 30%, preferably 50%, and most preferably 70% amino acid sequence identity between two or more of the RPS family members, RPS2, L6, N, or Prf. Examples of preferred conserved regions are shown (as boxed or designated sequences) in Figures 5 A and B, 6, 7, and 8 and include, without limitation, nucleotide binding site domains, leucine-rich repeats, leucine zipper domains, and P-loop domains.

By "detectably-labelled" is meant any means for marking and identifying the presence of a molecule, e.g., an oligonucleotide probe or primer, a gene or fragment thereof, or a cDNA molecule. Methods for detectably-labelling a molecule are well known in the art and include, without limitation, radioactive labelling (e.g., with an isotope such as <sup>32</sup>P or <sup>35</sup>S) and nonradioactive labelling (e.g., chemiluminescent labelling, e.g., fluorescein labelling).

By "biolistic transformation" is meant any method for introducing foreign molecules into a cell using velocity driven microprojectiles such as tungsten or gold particles. Such velocity-driven methods originate from pressure bursts which include, but are not limited to, helium-driven, air-driven, and gunpowder-driven techniques. Biolistic transformation may be applied to the transformation or transfection of a wide variety of cell types and intact tissues including, without



- 15 -

limitation, intracellular organelles (e.g., chloroplasts and mitochondria), bacteria, yeast, fungi, algae, pollen, animal tissue, plant tissue (e.g., leaf, seedling, embryo, epidermis, flower, meristem, and root), pollen, and cultured cells.

By "purified antibody" is meant antibody which is at least 60%, by weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably 90%, and most preferably at least 99%, by weight, antibody, e.g., an rps2-specific antibody. A purified rps antibody may be obtained, for example, by affinity chromatography using recombinantly-produced rps protein or conserved motif peptides and standard techniques.

By "specifically binds" is meant an antibody which recognizes and binds an rps protein but which does not substantially recognize and bind other molecules in a sample, e.g., a biological sample, which naturally includes rps protein.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

#### Detailed Description

The drawings will first be described.

#### Drawings

Figs. 1A - 1F are a schematic summary of the physical and RFLP analysis that led to the cloning of the RPS2 locus.

Fig. 1A is a diagram showing the alignment of the genetic and the RFLP maps of the relevant portion of *Arabidopsis thaliana* chromosome IV adapted from the map published by Lister and Dean (1993) Plant J. 4:745-750.

- 16 -

The RFLP marker L11F11 represents the left arm of the YUP11F11 YAC clone.

Fig. 1B is a diagram showing the alignment of relevant YACs around the *RPS2* locus. YAC constructs designated YUP16G5, YUP18G9 and YUP11F11 were provided by J. Ecker, University of Pennsylvania. YAC constructs designated EW3H7, EW11D4, EW11E4, and EW9C3 were provided by E. Ward, Ciba-Geigy, Inc.

Fig. 1C is a diagram showing the alignment of cosmid clones around the *RPS2* locus. Cosmid clones with the designation H are derivatives of the EW3H7 YAC clone whereas those with the designation E are derivatives of the EW11E4 YAC clone. Vertical arrows indicate the relative positions of RFLP markers between the ecotypes La-er and the *rps2-101N* plant. The RFLP markers were identified by screening a Southern blot containing more than 50 different restriction enzyme digests using either the entire part or pieces of the corresponding cosmid clones as probes. The cosmid clones described in Fig. 1C were provided by J. Giraudat, C.N.R.S., Gif-sur-Yvette, France.

Figs. 1D and 1E are maps of *EcoRI* restriction endonuclease sites in the cosmids E4-4 and E4-6, respectively. The recombination break points surrounding the *RPS2* locus are located within the 4.5 and 7.5 kb *EcoRI* restriction endonuclease fragments.

Fig. 1F is a diagram showing the approximate location of genes which encode the RNA transcripts which have been identified by polyA<sup>+</sup> RNA blot analysis. The sizes of the transcripts are given in kilobase pairs below each transcript.

Fig. 2 [SEQ ID NOS; 1-104, 196-201] is the complete nucleotide sequence of cDNA-4 comprising the *RPS2* gene locus. The three reading frames are shown below the nucleotide sequence. The deduced amino acid

- 17 -

sequence of reading frame "a" is provided and contains 909 amino acids. The methionine encoded by the ATG start codon is circled in open reading frame "a" of Fig. 2. The A of the ATG start codon is nucleotide 31 of Fig. 2.

5           Fig. 3 [SEQ ID NOS: 105-106] is the nucleotide sequence of the *avrRpt2* gene and its deduced amino acid sequence. A potential ribosome binding site is underlined. An inverted repeat is indicated by horizontal arrows at the 3' end of the open reading  
10 frame. The deduced amino acid sequence is provided below the nucleotide sequence of the open reading frame.

          Fig. 4 is a schematic summary of the complementation analysis that allowed functional confirmation that the DNA carried on p4104 and p4115  
15 (encoding cDNA-4) confers *RPS2* disease resistance activity to *Arabidopsis thaliana* plants previously lacking *RPS2* disease resistance activity. Small vertical marks along the "genome" line represent restriction enzyme *EcoRI* recognition sites, and the numbers above  
20 this line represent the size, in kilobase pairs (kb), of the resulting DNA fragments (see also Fig. 1E). Opposite "cDNAs" are the approximate locations of the coding sequences for RNA transcripts (See also Fig. 1F); arrowheads indicate the direction of transcription for  
25 cDNAs 4, 5, and 6. For functional complementation experiments, *rps2-201C/rps2-201C* plants were genetically transformed with the *Arabidopsis thaliana* genomic DNA sequences indicated; these sequences were carried on the named plasmids (derivatives of the binary cosmid vector  
30 pSLJ4541) and delivered to the plant via *Agrobacterium*-mediated transformation methods. The disease resistance phenotype of the resulting transformants following inoculation with *P. syringae* expressing *avrRpt2* is given as "Sus." (susceptible, no resistance response) or "Res."  
35 (disease resistant).

- 18 -

Fig. 5A [SEQ ID NOS: 107-136;, AND 142] shows regions of sequence similarity between the L-6 protein of flax, N protein of tobacco, Prf protein of tomato, and rps2 protein of *Arabidopsis*.

5 Fig. 5B [SEQ ID NOS: 107, 108, 137-140] shows sequence similarity between the N and L-6 proteins.

Fig. 6 [SEQ ID NOS: 141 and 142] shows a sequence analysis of RPS2 polypeptide showing polypeptide regions corresponding to an N-terminal hydrophobic region, a  
10 leucine zipper, NBSSs (kinase-1a, kinase-2, and kinase-3 motifs), and a predicted membrane integrated region.

Fig. 7 [SEQ ID NOS: 143-146] shows the amino acid sequence of the RPS2 LRR (amino acids 505-867). The top line indicates the consensus sequences for the RPS2 LRR.  
15 An "X" stands for an arbitrary amino acid sequence and an "a" stands for an aliphatic amino acid residue. The consensus sequence for the RPS2 LRR is closely related to the consensus for the yeast adenylate cyclase CYR1 LRR (PX Xa XXL XXL XXLXL XXNXaXXa). The amino acid residues  
20 that match the consensus sequence are shown in bold. Although this figure shows 14 LRRs, the C-terminal boundary of the LRR is not very clear because the LRR closer to the C-terminus does not fit the consensus sequence very well.

25 Fig. 8 [SEQ ID NO: 3] shows a sequence analysis of RPS2, indicating regions with similarity to leucine zipper, P-loop, membrane-spanning, and leucine-rich repeat motifs. Regions with similarity to defined functional domains are indicated with a line over the  
30 relevant amino acids. Potential N-glycosylation sequences are marked with a dot, and the location of the rps2-201 Thr to Pro mutation at amino acid 668 is marked with an asterisk.

Fig. 9 is a schematic representation of the  
35 transient assay method. The top panel shows the

- 19 -

essential principles of the assay. The bottom panel shows a schematic representation of the actual transient assay procedure. *Psp* NP53121 is used because it is a weak *Arabidopsis* pathogen, but potent in causing the HR when carrying an avirulence gene. In the absence of an HR, the damage to plant cells infected with NP53121 is minimal, enhancing the difference of GUS accumulation in cells that undergo the HR in comparison to those that do not. Prior to bombardment, one half of an *Arabidopsis* leaf is infiltrated with *P. syringae* (stippled side of leaf); the other half of the leaf serves as a noninfected control, an "internal" reference for the infected side, and as a measure of transformation efficiency.

Fig. 10, panels A-B, are photographs showing the complementation of the *rps2* mutant phenotype using the biolistic transient expression assay. The left sides of *rps2*-101C mutant leaves were infiltrated with *Psp* 3121/*avrRpt2*. Infiltrated leaves were cobombarded with either 35S-uidA plus  $\Delta$ GUS (Panel A) or 35S-uidA plus 35S-RPS2 (cDNA-2 clone 4) (Panel B). Note that in Panel B the infected side of the leaf shows less GUS activity than the uninfected side, indicating that the transformed cells on the infected side underwent an HR and that 35S-RPS2 complemented the mutant phenotype (see Fig. 9).

Fig. 11 is a schematic representation of pKEx4tr showing the structure of this cDNA expression vector. For convenience, the multiple cloning site contains the 8bp recognition sequences for *PmeI* and *NotI* and is flanked by T7 and T3 promoters. The region spanning the modified 35S promoter to the nopaline synthase 3' sequences (nos 3') was cloned into the *Hind III*-*EcoRI* site of pUC18, resulting in the loss of the *EcoRI* site.

Fig. 12 shows a nucleic acid sequence of the tomato *Prf* gene.

- 20 -

The Genetic Basis for Resistance to Pathogens

An overview of the interaction between a plant host and a microbial pathogen is presented. The invasion of a plant by a potential pathogen can have a range of outcomes delineated by the following outcomes: either the pathogen successfully proliferates in the host, causing associated disease symptoms, or its growth is halted by the host defenses. In some plant-pathogen interactions, the visible hallmark of an active defense response is the so-called hypersensitive response or "HR". The HR involves rapid necrosis of cells near the site of the infection and may include the formation of a visible dry brown lesion. Pathogens which elicit an HR on a given host are said to be avirulent on that host, the host is said to be resistant, and the plant-pathogen interaction is said to be incompatible. Strains which proliferate and cause disease on a particular host are said to be virulent; in this case the host is said to be susceptible, and the plant-pathogen interaction is said to be compatible.

"Classical" genetic analysis has been used successfully to help elucidate the genetic basis of plant-pathogen recognition for those cases in which a series of strains ( Races) of a particular fungal or bacterial pathogen are either virulent or avirulent on a series of cultivars (or different wild accessions) of a particular host species. In many such cases, genetic analysis of both the host and the pathogen revealed that many avirulent fungal and bacterial strains differ from virulent ones by the possession of one or more avirulence (avr) genes that have corresponding "resistance" genes in the host. This avirulence gene-resistance gene correspondence is termed the "gene-for-gene" model (Crute, et al., (1985) pp 197-309 in: *Mechanisms of Resistance to Plant Disease*. R.S.S. Fraser, ed.;

- 21 -

Ellingboe, (1981) Annu. Rev. Phytopathol. 19:125-143;  
Flor, (1971) Annu. Rev. Phytopathol. 9:275-296; Keen and  
Staskawicz, (1988) supra; and Keen et al. in: *Application  
of Biotechnology to Plant Pathogen Control*. I. Chet, ed.,  
5 John Wiley & Sons, 1993, pp. 65-88). According to a  
simple formulation of this model, plant resistance genes  
encode specific receptors for molecular signals generated  
by avr genes. Signal transduction pathway(s) then carry  
the signal to a set of target genes that initiate the HR  
10 and other host defenses (Gabriel and Rolfe, (1990) Annu.  
Rev. Phytopathol. 28:365-391). Despite this simple  
predictive model, the molecular basis of the avr-  
resistance gene interaction is still unknown.

One basic prediction of the gene-for-gene  
15 hypothesis has been convincingly confirmed at the  
molecular level by the cloning of a variety of bacterial  
avr genes (Innes, et al., (1993) J. Bacteriol. 175:4859-  
4869; Dong, et al., (1991) Plant Cell 3:61-72; Whelan et  
al., (1991) Plant Cell 3:49-59; Staskawicz et al., (1987)  
20 J. Bacteriol. 169:5789-5794; Gabriel et al., (1986)  
P.N.A.S., USA 83:6415-6419; Keen and Staskawicz, (1988)  
Annu. Rev. Microbiol. 42:421-440; Kobayashi et al.,  
(1990) Mol. Plant-Microbe Interact. 3:94-102 and (1990)  
Mol. Plant-Microbe Interact. 3:103-111). Many of these  
25 cloned avirulence genes have been shown to correspond to  
individual resistance genes in the cognate host plants  
and have been shown to confer an avirulent phenotype when  
transferred to an otherwise virulent strain. The avrRpt2  
locus was isolated from *Pseudomonas syringae* pv. tomato  
30 and sequenced by Innes et al. (Innes, R. et al. (1993) J.  
Bacteriol. 175:4859-4869). Fig. 3 is the nucleotide  
sequence and deduced amino acid sequence of the avrRpt2  
gene.

Examples of known signals to which plants respond  
35 when infected by pathogens include harpins from *Erwinia*

- 22 -

- (Wei et al. (1992) Science 257:85-88) and *Pseudomonas* (He et al. (1993) Cell 73:1255-1266); *avr4* (Joosten et al. (1994) Nature 367:384-386) and *avr9* peptides (van den Ackerveken et al (1992) Plant J. 2:359-366) from
- 5 *Cladosporium*; *PopA1* from *Pseudomonas* (Arlat et al. (1994) EMBO J. 13:543-553); *avrD*-generated lipopolysaccharide (Midland et al. (1993) J. Org. Chem. 58:2940-2945); and *NIP1* from *Rhynchosporium* (Hahn et al. (1993) Mol. Plant-Microbe Interact. 6:745-754).
- 10 Compared to *avr* genes, considerably less is known about plant resistance genes that correspond to specific *avr*-generated signals. The plant resistance gene, *RPS2* (*rps* for resistance to P*seudomonas* syringae), the first
- 15 disease resistance genes corresponds to a specific *avr* gene (*avrRpt2*). Some of the work leading up to the cloning of *RPS2* is described in Yu, et al., (1993), Molecular Plant-Microbe Interactions 6:434-443 and in Kunkel, et al., (1993) Plant Cell 5:865-875.
- 20 An apparently unrelated avirulence gene which corresponds specifically to plant disease resistance gene, *Pto*, has been isolated from tomato (*Lycopersicon esculentum*) (Martin et al., (1993) Science 262:1432-1436). Tomato plants expressing the *Pto* gene are resistant to
- 25 infection by strains of *Pseudomonas syringae* pv. *tomato* that express the *avrPto* avirulence gene. The amino acid sequence inferred from the *Pto* gene DNA sequence displays strong similarity to serine-threonine protein kinases, implicating *Pto* in signal transduction. No similarity to
- 30 the tomato *Pto* locus or any known protein kinases was observed for *RPS2*, suggesting that *RPS2* is representative of a new class of plant disease resistance genes.
- The isolation of a race-specific resistance gene from *Zea mays* (corn) known as *Hm1* has been reported
- 35 (Johal and Briggs (1992) Science 258:985-987). *Hm1*



- 23 -

confers resistance against specific races of the fungal pathogen *Cochliobolus carbonum* by controlling degradation of a fungal toxin, a strategy that is mechanistically distinct from the avirulence-gene specific resistance of the *RPS2*-*avrRpt2* resistance mechanism.

The cloned *RPS2* gene of the invention can be used to facilitate the construction of plants that are resistant to specific pathogens and to overcome the inability to transfer disease resistance genes between species using classical breeding techniques (Keen et al., (1993), supra). There now follows a description of the cloning and characterization of an *Arabidopsis thaliana* *RPS2* genetic locus, the *RPS2* genomic DNA, and the *RPS2* cDNA. The *avrRpt2* gene and the *RPS2* gene, as well as mutants *rps2-101C*, *rps2-102C*, and *rps2-201C* (also designated *rps2-201*), are described in Dong, et al., (1991) Plant Cell 3:61-72; Yu, et al., (1993) supra; Kunkel et al., (1993) supra; Whalen et al., (1991), supra; and Innes et al., (1993), supra). A mutant designated *rps2-101N* has also been isolated. The identification and cloning of the *RPS2* gene is described below.

#### *RPS2* Overcomes Sensitivity to Pathogens Carrying the *avrRpt2* Gene

To demonstrate the genetic relationship between an avirulence gene in the pathogen and a resistance gene in the host, it was necessary first to isolate an avirulence gene. By screening *Pseudomonas* strains that are known pathogens of crop plants related to *Arabidopsis*, highly virulent strains, *P. syringae* pv. *maculicola* (*Psm*) ES4326, *P. syringae* pv. *tomato* (*Pst*) DC3000, and an avirulent strain, *Pst* MM1065 were identified and analyzed as to their respective abilities to grow in wild type *Arabidopsis thaliana* plants (Dong et al., (1991) Plant

- 24 -

Cell, 3:61-72; Whalen et al., (1991) Plant Cell 3:49-59; MM1065 is designated JL1065 in Whalen et al.). *Psm* ES4326 or *Pst* DC3000 can multiply  $10^4$  fold in *Arabidopsis thaliana* leaves and cause water-soaked lesions that appear over the course of two days. *Pst* MM1065 multiplies a maximum of 10 fold in *Arabidopsis thaliana* leaves and causes the appearance of a mildly chlorotic dry lesion after 48 hours. Thus, disease resistance is associated with severely inhibited growth of the pathogen.

An avirulence gene (*avr*) of the *Pst* MM1065 strain was cloned using standard techniques as described in Dong et al. (1991), Plant Cell 3:61-72; Whalen et al., (1991) supra; and Innes et al., (1993), supra. The isolated avirulence gene from this strain was designated *avrRpt2*. Normally, the virulent strain *Psm* ES4326 or *Pst* DC3000 causes the appearance of disease symptoms after 48 hours as described above. In contrast, *Psm* ES4326/*avrRpt2* or *Pst* DC3000/*avrRpt2* elicits the appearance of a visible necrotic hypersensitivity response (HR) within 16 hours and multiplies 50 fold less than *Psm* ES4326 or *Pst* DC3000 in wild type *Arabidopsis thaliana* leaves (Dong et al., (1991), supra; and Whalen et al., (1991), supra). Thus, disease resistance in a wild type *Arabidopsis* plant requires, in part, an avirulence gene in the pathogen or a signal generated by the avirulence gene.

The isolation of four *Arabidopsis thaliana* disease resistance mutants has been described using the cloned *avrRpt2* gene to search for the host gene required for disease resistance to pathogens carrying the *avrRpt2* gene (Yu et al., (1993), supra; Kunkel et al., (1993), supra). The four *Arabidopsis thaliana* mutants failed to develop an HR when infiltrated with *Psm* ES4326/*avrRpt2* or *Pst* DC3000/*avrRpt2* as expected for plants having lost their disease resistance capacity. In the case of one of these

- 25 -

mutants, approximately 3000 five to six week old M<sub>2</sub> ecotype Columbia (Col-0 plants) plants generated by ethyl methanesulfonic acid (EMS) mutagenesis were hand-inoculated with *Psm* ES4326/*avrRpt2* and a single mutant, 5 *rps2-101C*, was identified (resistance to *Pseudomonas syringae*) (Yu et al., (1993), supra).

The second mutant was isolated using a procedure that specifically enriches for mutants unable to mount an HR (Yu et al., (1993), supra). When 10-day old 10 *Arabidopsis thaliana* seedlings growing on petri plates are infiltrated with *Pseudomonas syringae* pv. *phaseolicola* (*Psp*) NPS3121 versus *Psp* NPS3121/*avrRpt2*, about 90% of the plants infiltrated with *Psp* NPS3121 survive, whereas about 90%-95% of the plants infiltrated 15 with *Psp* NPS3121/*avrRpt2* die. Apparently, vacuum infiltration of an entire small *Arabidopsis thaliana* seedling with *Psp* NPS3121/*avrRpt2* elicits a systemic HR which usually kills the seedling. In contrast, seedlings infiltrated with *Psp* NPS3121 survive because *Psp* NPS3121 20 is a weak pathogen on *Arabidopsis thaliana*. The second disease resistance mutant was isolated by infiltrating 4000 EMS-mutagenized Columbia M<sub>2</sub> seedlings with *Psp* NPS3121/*avrRpt2*. Two hundred survivors were obtained. These were transplanted to soil and re-screened by hand 25 inoculation when the plants reached maturity. Of these 200 survivors, one plant failed to give an HR when hand-infiltrated with *Psm* ES4326/*avrRpt2*. This mutant was designated *rps2-102C* (Yu et al., (1993), supra).

A third mutant, *rps2-201C*, was isolated in a 30 screen of approximately 7500 M<sub>2</sub> plants derived from seed of *Arabidopsis thaliana* ecotype Col-0 that had been mutagenized with diepoxybutane (Kunkel et al., (1993), supra). Plants were inoculated by dipping entire leaf rosettes into a solution containing *Pst* DC3000/*avrRpt2* 35 bacteria and the surfactant Silwet L-77 (Whalen et al.,

- 26 -

(1991), supra), incubating plants in a controlled environment growth chamber for three to four days, and then visually observing disease symptom development. This screen revealed four mutant lines (carrying the  
5 *rps2-201C*, *rps2-202C*, *rps2-203C*, and *rps2-204C* alleles), and plants homozygous for *rps2-201C* were a primary subject for further study (Kunkel et al., (1993), supra and the instant application).

Isolation of the fourth *rps2* mutant, *rps2-101N*,  
10 has not yet been published. This fourth isolate is either a mutant or a susceptible *Arabidopsis* ecotype. Seeds of the *Arabidopsis* Nossen ecotype were gamma-irradiated and then sown densely in flats and allowed to germinate and grow through a nylon mesh. When the plants  
15 were five to six weeks old, the flats were inverted, the plants were partially submerged in a tray containing a culture of *Psm* ES4326/*avrRpt2*, and the plants were vacuum infiltrated in a vacuum desiccator. Plants inoculated this way develop an HR within 24 hours. Using this  
20 procedure, approximately 40,000 plants were screened and one susceptible plant was identified. Subsequent RFLP analysis of this plant suggested that it may not be a Nossen mutant but rather a different *Arabidopsis* ecotype that is susceptible to *Psm* ES4326/*avrRpt2*. This plant is  
25 referred to as *rps2-101N*. The isolated mutants *rps2-101C*, *rps2-102C*, *rps2-201C*, and *rps2-101N* are referred to collectively as the "*rps2* mutants".

The *rps2* Mutants Fail to Specifically Respond to the Cloned Avirulence Gene, *avrRpt2*

30 The *RPS2* gene product is specifically required for resistance to pathogens carrying the avirulence gene, *avrRpt2*. A mutation in *Rps2* polypeptide that eliminates or reduces its function would be observable as the absence of a hypersensitive response upon infiltration of

- 27 -

the pathogen. The *rps2* mutants displayed disease symptoms or a null response when infiltrated with *Psm* ES4326/*avrRpt2*, *Pst* DC3000/*avrRpt2* or *Psp* NPS3121/*avrRpt2*, respectively. Specifically, no HR response was elicited, indicating that the plants were susceptible and had lost resistance to the pathogen despite the presence of the *avrRpt2* gene in the pathogen.

Pathogen growth in *rps2* mutant plant leaves was similar in the presence and absence of the *avrRpt2* gene. *Psm* ES4326 and *Psm* ES4326/*avrRpt2* growth in *rps2* mutants was compared and found to multiply equally well in the *rps2* mutants, at the same rate that *Psm* ES4326 multiplied in wild-type *Arabidopsis* leaves. Similar results were observed for *Pst* DC3000 and *Pst* DC3000/*avrRpt2* growth in *rps2* mutants.

The *rps2* mutants displayed a HR when infiltrated with *Pseudomonas* pathogens carrying other *avr* genes, *Psm* ES4326/*avrB*, *Pst* DC3000/*avrB*, *Psm* ES4326/*avrRpm1*, *Pst* DC3000/*avrRpm1*. The ability to mount an HR to an *avr* gene other than *avrRpt2* indicates that the *rps2* mutants isolated by selection with *avrRpt2* are specific to *avrRpt2*.

#### Mapping and Cloning of the RPS2 Gene

Genetic analysis of *rps2* mutants *rps2-101C*, *rps2-102C*, *rps-201C* and *rps-101N* showed that they all corresponded to genes that segregated as expected for a single Mendelian locus and that all four were most likely allelic. The four *rps2* mutants were mapped to the bottom of chromosome IV using standard RFLP mapping procedures including polymerase chain reaction (PCR)-based markers (Yu et al., (1993), *supra*; Kunkel et al., (1993), *supra*; and Mindrinos, M., unpublished). Segregation analysis showed that *rps2-101C* and *rps2-102C* are tightly linked to the PCR marker, PG11, while the RFLP marker M600 was used

- 28 -

to define the chromosome location of the *rps2-201C* mutation (Fig. 1A) (Yu et al., (1993), supra; Kunkel et al., (1993), supra). *RPS2* has subsequently been mapped to the centromeric side of PG11.

- 5 Heterozygous *RPS2/rps2* plants display a defense response that is intermediate between those displayed by the wild-type and homozygous *rps2/rps2* mutant plants (Yu, et al., (1993), supra; and Kunkel et al., (1993), supra). The heterozygous plants mounted an HR in response to *Psm*  
10 ES4326/*avrRpt2* or *Pst* DC3000/*avrRpt2* infiltration; however, the HR appeared later than in wild type plants and required a higher minimum inoculum (Yu, et al., (1993), supra; and Kunkel et al., (1993), supra).

#### High Resolution Mapping of the *RPS2* Gene and *RPS2* cDNA

##### 15 Isolation

- To carry out map-based cloning of the *RPS2* gene, *rps2-101N/rps2-101N* was crossed with Landsberg erecta *RPS2/RPS2*. Plants of the  $F_1$  generation were allowed to self pollinate (to "self") and 165  $F_2$  plants were selfed  
20 to generate  $F_3$  families. Standard RFLP mapping procedures showed that *rps2-101N* maps close to and on the centromeric side of the RFLP marker, PG11. To obtain a more detailed map position, *rps2-101N/rps2-101N* was crossed with a doubly marked Landsberg erecta strain  
25 containing the recessive mutations, *cer2* and *ap2*. The genetic distance between *cer2* and *ap2* is approximately 15 cM, and the *rps2* locus is located within this interval.  $F_2$  plants that displayed either a *CER2 ap2* or a *cer2 AP2* genotype were collected, selfed, and scored for *RPS2* by  
30 inoculating at least 20  $F_3$  plants for each  $F_2$  with *Psm* ES4326/*avrRpt2*. DNA was also prepared from a pool of approximately 20  $F_3$  plants for each  $F_2$  line. The *CER2 ap2* and *cer2 AP2* recombinants were used to carry out a chromosome walk that is illustrated in Figure 1.

- 29 -

As shown in Figure 1, *RPS2* was mapped to a 28-35 kb region spanned by cosmid clones E4-4 and E4-6. This region contains at least six genes that produce detectable transcripts. There were no significant differences in the sizes of the transcripts or their level of expression in the *rps2* mutants as determined by RNA blot analysis. cDNA clones of each of these transcripts were isolated and five of these were sequenced. As is described below, one of these transcripts, cDNA-4, was shown to correspond to the *RPS2* locus. From this study, three independent cDNA clones (cDNA-4-4, cDNA-4-5, and cDNA-4-11) were obtained corresponding to *RPS2* from Columbia ecotype wild type plants. The apparent sizes of *RPS2* transcripts were 3.8 and 3.1 kb as determined by RNA blot analysis.

A fourth independent cDNA-4 clone (cDNA-4-2453) was obtained using map-based isolation of *RPS2* in a separate study. Yeast artificial chromosome (YAC) clones were identified that carry contiguous, overlapping inserts of *Arabidopsis thaliana* ecotype Col-0 genomic DNA from the M600 region spanning approximately 900 kb in the *RPS2* region. *Arabidopsis* YAC libraries were obtained from J. Ecker and E. Ward, supra and from E. Grill (Grill and Somerville (1991) Mol. Gen. Genet. 226:484-490). Cosmids designated "H" and "E" were derived from the YAC inserts and were used in the isolation of *RPS2* (Fig. 1).

The genetic and physical location of *RPS2* was more precisely defined using physically mapped RFLP, RAPD (random amplified polymorphic DNA) and CAPS (cleaved amplified polymorphic sequence) markers. Segregating populations from crosses between plants of genotype *RPS2/RPS2* (No-0 wild type) and *rps2-201/rps2-201* (Col-0 background) were used for genetic mapping. The *RPS2* locus was mapped using markers 17B7LE, PG11, M600 and other markers. For high-resolution genetic mapping, a

- 30 -

set of tightly linked RFLP markers was generated using insert end fragments from YAC and cosmid clones (Fig. 1) (Kunkel et al. (1993), supra; Konieczny and Ausubel (1993) Plant J. 4:403-410; and Chang et al. (1988) PNAS USA 85:6856-6860). Cosmid clones E4-4 and E4-6 were then used to identify expressed transcripts (designated cDNA-4, -5, -6, -7, -8 of Fig 1F) from this region, including the cDNA-4-2453 clone.

#### RPS2 DNA Sequence Analysis

10 DNA sequence analysis of cDNA-4 from wild-type Col-0 plants and from mutants *rps2-101C*, *rps2-102C*, *rps2-201C* and *rps2-101N* showed that cDNA-4 corresponds to *RPS2*. DNA sequence analysis of *rps2-101C*, *rps2-102C* and *rps2-201C* revealed changes from the wild-type sequence as  
15 shown in Table 1. The numbering system in Table 1 starts at the ATG start codon encoding the first methionine where A is nucleotide 1. DNA sequence analysis of cDNA-4 corresponding to mutant *rps2-102C* showed that it differed from the wild type sequence at amino acid residue 476.  
20 Moreover, DNA sequence analysis of the cDNA corresponding to cDNA-4 from *rps2-101N* showed that it contained a 10 bp insertion at amino acid residue 581, a site within the leucine-rich repeat region which causes a shift in the *RPS2* reading frame. Mutant *rps2-101C* contains a mutation  
25 that leads to the formation of a chain termination codon. The DNA sequence of mutant allele *rps2-201C* revealed a mutation altering a single amino acid within a segment of the LRR region that also has similarity to the helix-loop-helix motif, further supporting the designation of  
30 this locus as the *RPS2* gene. The DNA and amino acid sequences are shown in Figure 2.



- 31 -

Table 1

	Mutant	Wild type	position of mutation	Change
	<i>rps2-101C</i>	703 TGA 705	704	TAA Stop Codon
5	<i>rps2-101N</i>	1741 GTG 1743	1741	GTGGAGTTGTATG Insertion
	<i>rps2-102C</i>	1426 AGA 1428 arg	1427	AAA Amino acid 476 lys
10	<i>rps2-201C</i>	2002 ACC 2004 thr	2002	CCC Amino acid pro

DNA sequence analysis of cDNA-4 corresponding to *RPS2* from wild-type Col-0 plants revealed an open reading frame (between two stop codons) spanning 2,751 bp. There are 2,727 bp between the first methionine codon of this  
 15 reading frame and the 3'-stop codon, which corresponds to a deduced 909 amino acid polypeptide (See open reading frame "a" of Fig. 2). The amino acid sequence has a relative molecular weight of 104,460 and a pI of 6.51.

As discussed below, *RPS2* belongs to a new class of  
 20 disease resistance genes; the structure of the *Rps2* polypeptide does not resemble the protein structure of the product of the previously cloned and publicized avirulence gene-specific plant disease resistance gene, *Pto*, which has a putative protein kinase domain. From  
 25 the above analysis of the deduced amino acid sequence, *RPS2* contains several distinct protein domains conserved in other proteins from both eukaryotes and prokaryotes. These domains include, but are not limited, to Leucine Rich Repeats (LRR) (Kobe and Deisenhofer, (1994) Nature  
 30 366:751-756); nucleotide binding site, e.g. the kinase 1a motif (P-loop) (Saraste et al. (1990) Trends in Biological Sciences TIBS 15:430-434; Helix-Loop-Helix (Murre et al. (1989) Cell 56:777-783; and Leucine Zipper

- 32 -

(Rodrigues and Park (1993) Mol. Cell Biol. 13:6711-6722). The amino acid sequence of Rps2 contains a LRR motif (LRR motif from amino acid residue 505 to amino acid residue 867), which is present in many known proteins and which is thought to be involved in protein-protein interactions and may thus allow interaction with other proteins that are involved in plant disease resistance. The N-terminal portion of the Rps2 polypeptide LRR is, for example, related to the LRR of yeast (*Saccharomyces cerevisiae*) adenylate cyclase, CYR1. A region predicted to be a transmembrane spanning domain (Klein et al. (1985) Biochim., Biophys. Acta 815:468-476) is located from amino acid residue 350 to amino acid residue 365, N-terminal to the LRR. An ATP/GTP binding site motif (P-loop) is predicted to be located between amino acid residue 177 and amino acid residue 194, inclusive. The motifs are discussed in more detail below.

From the above analysis of the deduced amino acid sequence, the Rps2 polypeptide may have a membrane-receptor structure which consists of an N-terminal extracellular region and a C-terminal cytoplasmic region. Alternatively, the topology of the Rps2 may be the opposite: an N-terminal cytoplasmic region and a C-terminal extracellular region. LRR motifs are extracellular in many cases and the Rps2 LRR contains five potential N-glycosylation sites.

#### Identification of RPS2 by Functional Complementation

Complementation of *rps2-201* homozygotes with genomic DNA corresponding to *Arabidopsis thaliana* functionally confirmed that the genomic region encoding cDNA-4 carries RPS2 activity. Cosmids were constructed that contained overlapping contiguous sequences of wild type *Arabidopsis thaliana* DNA from the RPS2 region contained in YACs EW11D4, EW9C3, and YUP11F1 of Fig. 1

- 33 -

and Fig. 4. The cosmid vectors were constructed from pSLJ4541 (obtained from J. Jones, Sainsbury Institute, Norwich, England) which contains sequences that allow the inserted sequence to be integrated into the plant genome via *Agrobacterium*-mediated transformation (designated "binary cosmid"). "H" and "E" cosmids (Fig. 1) were used to identify clones carrying DNA from the *Arabidopsis thaliana* genomic *RPS2* region.

More than forty binary cosmids containing inserted *RPS2* region DNA were used to transform *rps2-201* homozygous mutants utilizing *Agrobacterium*-mediated transformation (Chang et al. ((1990) p. 28, Abstracts of the Fourth International Conference on *Arabidopsis* Research, Vienna, Austria). Transformants which remained susceptible (determined by methods including the observed absence of an HR following infection to *P. syringae* pv. *phaseolicola* strain 3121 carrying *avrRpt2* and Psp 3121 without *avrRpt2*) indicated that the inserted DNA did not contain functional *RPS2*. These cosmids conferred the "Sus." or susceptible phenotype indicated in Fig. 4. Transformants which had acquired *avrRpt2*-specific disease resistance (determined by methods including the display of a strong hypersensitive response (HR) when inoculated with Psp 3121 with *avrRpt2*, but not following inoculation with Psp 3121 without *avrRpt2*) suggested that the inserted DNA contained a functional *RPS2* gene capable of conferring the "Res." or resistant phenotype indicated in Fig. 4. Transformants obtained using the pD4 binary cosmid displayed a strong resistance phenotype as described above. The presence of the insert DNA in the transformants was confirmed by classical genetic analysis (the tight genetic linkage of the disease resistance phenotype and the kanamycin resistance phenotype conferred by the cotransformed selectable marker) and Southern analysis. These results indicated that *RPS2* is

- 34 -

encoded by a segment of the 18 kb *Arabidopsis thaliana* genomic region carried on cosmid pD4 (Fig. 4).

To further localize the *RPS2* locus and confirm its ability to confer a resistance phenotype on the *rps2-201* homozygous mutants, a set of six binary cosmids containing partially overlapping genomic DNA inserts were tested. The overlapping inserts pD2, pD4, pD14, pD15, pD27, and pD47 were chosen based on the location of the transcription corresponding to the five cDNA clones in the *RPS2* region (Fig. 4). These transformation experiments utilized a vacuum infiltration procedure (Bechtold et al. (1993) C.R. Acad. Sci. Paris 316:1194-1199) for *Agrobacterium*-mediated transformation. *Agrobacterium*-mediated transformations with cosmids pD2, pD14, pD15, pD39, and pD46 were performed using a root transformation/regeneration protocol (Valveekens et al. (1988), PNAS 85:5536-5540). The results of pathogen inoculation experiments assaying for *RPS2* activity in these transformants is indicated in Fig. 4.

These experiments were further confirmed using a modification of the vacuum filtration procedure. In particular, the procedure of Bechtold et al. (*supra*) was modified such that plants were grown in peat-based potting soil covered with a screen, primary inflorescences were removed, and plants with secondary inflorescences (approximately 3 to 15 cm in length) were inverted directly into infiltration medium, infiltrated, and then grown to seed harvest without removal from soil (detailed protocol available on the AAtDB computer database (43). The presence of introduced sequences in the initial pD4 transformant was verified by DNA blot analysis with a pD4 vector and insert sequences (separately) as probes. The presence of the expected sequences in transformants obtained with the vacuum infiltration protocol was also confirmed by DNA blot

- 35 -

analysis. Root transformation experiments (19) were performed with an easily regenerable *rps2-201/rps2-201* x No-0 mapping population. Transformants were obtained for pD4 with in plant transformation, for pD2, 14, 16, 39, 5 and 49 with root transformation, and for pD2, 4, 14, 15, 27, and 47 with vacuum infiltration as modified.

Additional transformation experiments utilized binary cosmids carrying the complete coding region and more than 1 kb of upstream genomic sequence for only 10 cDNA-4 or cDNA-6. Using the vacuum infiltration transformation method, three independent transformants were obtained that carried the wild-type cDNA-6 genomic region in a *rps2-201c* homozygous background (pAD431 of Fig. 4). None of these plants displayed *avrRpt2*- 15 dependent disease resistance. Homozygous *rps2-201c* mutants were transformed with wild-type genomic cDNA-4 (p4104 and p4115, each carrying Col-0 genomic sequences corresponding to all of the cDNA-4 open reading frame, plus approximately 1.7 kb of 5' upstream sequence and 20 approximately 0.3 kb of 3' sequence downstream of the stop codon). These p4104 and p4115 transformants displayed a disease resistance phenotype similar to the wild-type *RPS2* homozygotes from which the *rps2* were derived. Additional mutants (*rps2-101N* and *rps2-101C* 25 homozygotes) also displayed *avrRpt2*-dependent resistance when transformed with the cDNA-4 genomic region.

#### RPS2 Sequences Allow Detection of Other Resistance Genes

DNA blot analysis of *Arabidopsis thaliana* genomic DNA using *RPS2* cDNA as the probe showed that *Arabidopsis* 30 contains several DNA sequences that hybridize to *RPS2* or a portion thereof, suggesting that there are several related genes in the *Arabidopsis* genome.

From the aforementioned description and the nucleic acid sequence shown in Fig. 2, it is possible to

- 36 -

isolate other plant disease resistance genes having about 50% or greater sequence identity to the *RPS2* gene. Detection and isolation can be carried out with an oligonucleotide probe containing the *RPS2* gene or a portion thereof greater than 9 nucleic acids in length, and preferably greater than about 18 nucleic acids in length. Probes to sequences encoding specific structural features of the Rps2 polypeptide are preferred as they provide a means of isolating disease resistance genes having similar structural domains. Hybridization can be done using standard techniques such as are described in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, (1989).

For example, high stringency conditions for detecting the *RPS2* gene include hybridization at about 42°C, and about 50% formamide; a first wash at about 65°C, about 2X SSC, and 1% SDS; followed by a second wash at about 65°C and about 0.1% x SSC. Lower stringency conditions for detecting *RPS* genes having about 50% sequence identity to the *RPS2* gene are detected by, for example, hybridization at about 42°C in the absence of formamide; a first wash at about 42°C, about 6X SSC, and about 1% SDS; and a second wash at about 50°C, about 6X SSC, and about 1% SDS. An approximately 350 nucleotide DNA probe encoding the middle portion of the LRR region of Rps2 was used as a probe in the above example. Under lower stringency conditions, a minimum of 5 DNA bands were detected in *Bam*HI digested *Arabidopsis thaliana* genomic DNA as sequences having sufficient sequence identity to hybridize to DNA encoding the middle portion of the LRR motif of Rps2. Similar results were obtained using a probe containing a 300 nucleotide portion of the *RPS2* gene encoding the extreme N-terminus of Rps2 outside of the LRR motif.

- 37 -

Isolation of other disease resistance genes is performed by PCR amplification techniques well known to those skilled in the art of molecular biology using oligonucleotide primers designed to amplify only sequences flanked by the oligonucleotides in genes having sequence identity to *RPS2*. The primers are optionally designed to allow cloning of the amplified product into a suitable vector.

#### The RPS Disease-Resistance Gene Family

As discussed above, we have discovered that the *Arabidopsis RPS2* gene described herein is representative of a new class of plant resistance genes. Analysis of the derived amino acid sequence for *RPS2* revealed several regions of similarity with known polypeptide motifs (see, e.g., Schneider et al., Genes Dev. 6:797 (1991)). Most prominent among these is a region of multiple, leucine-rich repeats (LRRs). The LRR motif has been implicated in protein-protein interactions and ligand binding in a diverse array of proteins (see, e.g., Kornfield et al., Annu. Rev. Biochem. 64:631 (1985); Alber, Curr. Opin. Gen. Dev. 2:205 (1992); Lupas et al., Science 252:1162 (1991); Saraste et al., Trend Biochem. Sci. 15:430 (1990)). In one example, LRRs form the hormone binding sites of mammalian gonadotropin hormone receptors (see, e.g., Lupas et al., Science 252:1162 (1991)) and, in another example, a domain of yeast adenylate cyclase that interacts with the RAS2 protein (Kornfield et al., Annu. Rev. Biochem. 64:631 (1985)). In *RPS2*, the LRR domain spans amino acids 503-867 and contains fourteen repeat units of length 22-26 amino acids. A portion of each repeat resembles the LRR consensus sequence (I/L/V)XXLXXLXX(I/L)XL. In Figure 7, the LRRs from *RPS2* are shown, as well as an *RPS2* consensus sequence. Within the *RPS2* LRR region, five (of six) sequences matching the

- 38 -

N-glycosylation consensus sequence [NX(S/T)] were observed (Figure 8, marked with a dot). In particular, N-glycosylation is predicted to occur at amino acids 158, 543, 666, 757, 778, 787. Interestingly, the single  
5 nucleotide difference between functional *RPS2* and mutant allele *rps2-201* is within the LRR coding region, and this mutation disrupts one of the potential glycosylation sites.

Also observed in the deduced amino acid sequence  
10 for *RPS2* is a second potential protein-protein interaction domain, a leucine zipper (see, e.g., von Heijne, J. Mol. Biol. 225:487 (1992)), at amino acids 30-57. This region contains four contiguous heptad repeats that match the leucine zipper consensus sequence  
15 (I/R)XDLXXX. Leucine zippers facilitate the dimerization of transcription factors by formation of coiled-coil structures, but no sequences suggestive of an adjacent DNA binding domain (such as a strongly basic region or a potential zinc-finger) were detected in *RPS2*. Coiled-  
20 coil regions also promote specific interactions between proteins that are not transcription factors (see, e.g., Ward et al., Plant Mol. Biol. 14:561 (1990); Ecker, Methods 1:186 (1990); Grill et al., Mol. Gen. Genet. 226:484 (1991)), and computer database similarity  
25 searches with the region spanning amino acids 30-57 of *RPS2* revealed highest similarity to the coiled-coil regions of numerous myosin and paramyosin proteins.

A third *RPS2* motif was found at the sequence GPGGVGKT at deduced amino acids 182-189. This portion of  
30 *RPS2* precisely matches the generalized consensus for the phosphate-binding loop (P-loop) of numerous ATP- and GTP-binding proteins (see, e.g., Saraste et al., supra). The postulated *RPS2* P-loop is similar to those found in RAS proteins and ATP synthase  $\beta$ -subunits (Saraste et al.,  
35 supra), but surprisingly is most similar to the published



- 39 -

P-loop sequences for the *nifH* and *chvD* genes, respectively. The presence of this P-loop sequence strongly suggests nucleotide triphosphate binding as one aspect of *RPS2* function. This domain is also referred to  
5 as a kinase-1a motif (or a nucleotide binding site, or NBS). Other conserved NBSs are present in the *RPS2* sequence; these NBSs include a kinase-2 motif at amino acids 258-262 and a kinase-3a motif at amino acids 330-335.

10 Finally, inspection of the *RPS2* sequence reveals a fourth *RPS2* motif, a potential membrane-spanning domain located at amino acids 340-360. Within this region, a conserved GLPLAL motif is found at amino acids 347-352. The presence of the membrane-spanning domain raises the  
15 possibility that the *RPS2* protein is membrane localized, with the N-terminal leucine zipper and P-loop domains residing together on the opposite side of the membrane from the LRR region. An orientation in which the C-terminal LRR domain is extracellular is suggested by the  
20 fact that five of the six potential N-linked glycosylation sites occur C-terminal to the proposed membrane-spanning domain, as well as by the overall more positive charge of the N-terminal amino acid residues (see, e.g., Kornfield et al., supra; von Heijne, supra).  
25 A number of proteins that contain LRRs are postulated or known to be membrane-spanning receptors in which the LRRs are displayed extracellularly as a ligand-binding domain (see, e.g., Lopez et al., Proc. Natl. Acad. Sci. 84:5615 (1987); Braun et al., EMBO J. 10:1885 (1991); Schneider  
30 et al., supra).

The plant kingdom contains hundreds of resistance genes that are necessarily divergent since they control different resistance specificities. However, plant defense responses such as production of activated oxygen  
35 species, PR-protein gene expression, and the

- 40 -

hypersensitive response are common to diverse plant-pathogen interactions. This implies that there are points of convergence in the defense signal transduction pathways downstream of initial pathogen recognition, and also suggests that similar functional motifs may exist among diverse resistance gene products. Indeed, *RPS2* is dissimilar from previously described disease resistance genes such as *Hm1* or *Pto* (see, e.g., Johal et al., supra; Martin et al., supra), and thus represents a new class of genes having disease resistance capabilities.

Isolation of Other Members of the RPS Disease-Resistance Gene Family Using Conserved Motif Probes and Primers

We have discovered that the *RPS2* motifs described above are conserved in other disease-resistance genes, including, without limitation, the N protein, the L6 protein, and the Prf protein. As shown in Fig. 5(A and B), we have determined that the L6 polypeptide of flax, the N polypeptide of tobacco, and the Prf polypeptide of tomato each share unique regions of similarity (including, but not limited to, the leucine-rich repeats, the membrane-spanning domain, the leucine zipper, and the P-loop and other NBS domains).

On the basis of this discovery, the isolation of virtually any member of the RPS gene family is made possible using standard techniques. In particular, using all or a portion of the amino acid sequence of a conserved RPS motif (for example, the amino acid sequences defining any RPS P-loop, NBS, leucine-rich repeat, leucine zipper, or membrane-spanning region), one may readily design RPS oligonucleotide probes, including RPS degenerate oligonucleotide probes (i.e., a mixture of all possible coding sequences for a given amino acid sequence). These oligonucleotides may be based upon the sequence of either strand of the DNA comprising the

- 41 -

- motif. General methods for designing and preparing such probes are provided, for example, in Ausubel et al., supra and *Guide to Molecular Cloning Techniques*, 1987, S. L. Berger and A. R. Kimmel, eds., Academic Press, New York.
- 5 These oligonucleotides are useful for RPS gene isolation, either through their use as probes capable of hybridizing to RPS complementary sequences or as primers for various polymerase chain reaction (PCR) cloning strategies.
- 10 Hybridization techniques and procedures are well known to those skilled in the art and are described, for example, in Ausubel et al., supra and *Guide to Molecular Cloning Techniques*, 1987, S. L. Berger and A. R. Kimmel, eds., Academic Press, New York. If desired, a
- 15 combination of different oligonucleotide probes may be used for the screening of the recombinant DNA library. The oligonucleotides are labelled with  $^{32}\text{P}$  using methods known in the art, and the detectably-labelled oligonucleotides are used to probe filter replicas from a
- 20 recombinant plant DNA library. Recombinant DNA libraries may be prepared according to methods well known in the art, for example, as described in Ausubel et al., supra. Positive clones may, if desired, be rescreened with additional oligonucleotide probes based upon other RPS
- 25 conserved regions. For example, an RPS clone identified based on hybridization with a P-loop-derived probe may be confirmed by re-screening with a leucine-rich repeat-derived oligonucleotide.

- As discussed above, RPS oligonucleotides may also
- 30 be used as primers in PCR cloning strategies. Such PCR methods are well known in the art and described, for example, in *PCR Technology*, H.A. Erlich, ed., Stockton Press, London, 1989; *PCR Protocols: A Guide to Methods and Applications*, M.A. Innis, D.H. Gelfand, J.J. Sninsky,
- 35 and T.J. White, eds., Academic Press, Inc., New York,

- 42 -

1990; and Ausubel et al., supra. If desired, members of the RPS disease-resistance gene family may be isolated using the PCR "RACE" technique, or Rapid Amplification of cDNA Ends (see, e.g., Innis et al., supra). By this method, oligonucleotide primers based on an RPS conserved domain are oriented in the 3' and 5' directions and are used to generate overlapping PCR fragments. These overlapping 3'- and 5'-end RACE products are combined to produce an intact full-length cDNA. This method is described in Innis et al., supra; and Frohman et al., Proc. Natl. Acad. Sci. 85:8998, 1988.

Any number of probes and primers according to the invention may be designed based on the conserved RPS motifs described herein. Preferred motifs are boxed in the sequences shown in Fig. 5(A or B). In particular, oligonucleotides according to the invention may be based on the conserved P-loop domain, the amino acids of which are shown below:

MOTIF 1

20	L6	G MGGIGKTTTA [SEQ ID NO: 110]
	N	G MGGVGKTTIA [SEQ ID NO: 111]
	PrfP	G MPGLGKTTLA [SEQ ID NO: 112]
	RPS2	G PGGVGKTTLM [SEQ ID NO: 113]

From these sequences, appropriate oligonucleotides are designed and prepared using standard methods. Particular examples of RPS oligonucleotides based on the P-loop domain are as follows (N is A, C, T, or G).

Based on MOTIF 1:

30	5'	GGNATGGGNGGNNTNGGNAA(A or G)ACNAC 3' [SEQ ID NO: 158]
	5'	NCGNG(A/T)NGTNA(T/G)(G/A/T)A(T/A)NCGNA 3' [SEQ ID NO: 159]
	5'	GG(T or A)NT(T or G or C)GG(T or A)AA(G or A)AC(T or C or A)AC 3' [SEQ ID NO: 160]

- 43 -

5' GGNATGGGNGGNNTNGGNAA(A or G)ACNAC 3' [SEQ ID NO: 158]

5' N(G or A)(C or T)N(A or G)(A or G or T)NGTNGT(C or T)TTNCCNANNCCN(G or L)(G or C)N(G or A)(T or G)NCC 3' [SEQ ID NO: 161]

5' GGN(C or A)(T or C)N(G or C)(G or C)NGGNNTNGGNAA (A or G)ACNAC 3' [SEQ ID NO: 162]

Other conserved RPS motifs useful for  
10 oligonucleotide design are shown below. These motifs are  
also depicted in the sequence of Fig. 5(A or B).

MOTIF 2

	L6	FKILVV LDDVD [SEQ ID NO: 114]
	N	KKVLIV LDDID [SEQ ID NO: 115]
15	PrfP	KRFLIL IDDVW [SEQ ID NO: 116]
	RPS2	KRFLLL LDDVW [SEQ ID NO: 117]

MOTIF 3

	L6	SRFIIT SR [SEQ ID NO: 118]
	N	SRIIIT TR [SEQ ID NO: 119]
20	PrfP	SRIILT TR [SEQ ID NO: 120]
	RPS2	CKVMFT TR [SEQ ID NO: 121]

MOTIF 4

	L6	GLPLTLK V [SEQ ID NO: 122]
	N	GLPLALK V [SEQ ID NO: 123]
25	PrfP	GLPLSVV L [SEQ ID NO: 124]
	RPS2	GLPLALI T [SEQ ID NO: 125]

MOTIF 5

	L6	KISYDAL [SEQ ID NO: 126]
	N	KISYDGL [SEQ ID NO: 127]
30	PrfP	GFSYKNL [SEQ ID NO: 128]
	RPS2	KFSYDNL [SEQ ID NO: 129]

- 44 -

From the above motifs and the sequence motifs designated in Figure 5A and B, appropriate oligonucleotides are designed and prepared. Particular examples of such RPS oligonucleotides are as follows (N is A, T, C, or G).

- 5            Based on MOTIF 2:
- 5'    T(T or C)GA(T or C)GA(T or C)(A or G)T(T or G  
or C)(T or G)(A or G)(T or G or C)(G or A)A  
3' [SEQ ID NO: 163]
- 10          5'    T(T or C)CCA(G or C or A)A(T or C)(G or  
A)TC(A or G)TCNA 3' [SEQ ID NO: 164]
- 5'    (C or G or A)(T or C)(C or A)NA(T or C)(G or  
A)TC(G or A)TCNA(G or A or T)NA(G or A or  
C)NANNA(G or A)NA 3' [SEQ ID NO: 165]
- 15          5'    (T or A)(T or A)N(A or C)(A or G)(A or G)(T  
or G or A)TN(T or C)TNNTN(G or T or C)TN(A or  
T or C)TNGA(T or C)GA 3' [SEQ ID NO: 166]
- Based on MOTIF 3:
- 5'    NCGNG(A or T)NGTNA(T or G)(G or A or T)A(T or  
A)NCGNGA 3' [SEQ ID NO: 167]
- 20          5'    NCGNG(A or T)NGTNA(T or G)(G or A or T)A(T or  
A)NCGNGA 3' [SEQ ID NO: 167]
- 5'    NC(G or T)N(G or C)(A or T)NGTNA(A or G or  
T)(A or G or T)AT(A or G or T)AATNG 3' [SEQ ID  
NO: 168].
- 25           Based on MOTIF 4:
- 5'    NA(G or A)NGGNA(G or A)NCC 3' [SEQ ID NO: 169]
- 5'    GG(T or A)(T or C)T(T or G or C)CC(T or A)(T  
or C)T(T or G or C)GC(T or C or A)(T or C)T  
3' [SEQ ID NO: 170]
- 30          5'    A(A or G)(T or G or A)GC(G or C or A)A(G or  
A)(T or A)GG(G or C or A)A(G or A)(A or G or  
T or C)C C 3' [SEQ ID NO: 171]
- 5'    NA(G or A)NGGNA(G or A)NCC 3' [SEQ ID NO:  
169]
- 35          5'    N(A or G)NN(T or A)(T or C)NA(G or C or A)N(C  
or G)(A or T or C)NA(G or A)NGGNA(G or A)NCC  
3' [SEQ ID NO: 172]

- 45 -

5' GGN(T or C)TNCCN(T or C)TN(G or A or T)(C or G)N(T or G or C)T 3' [SEQ ID NO: 173]

Based on MOTIF 5:

5 5' A(A or G)(A or G)TT(A or G)TC(A or G)TA(G or A or T)(G or C)(T or A)(G or A)A(T or A)(C or T)TT 3' [SEQ ID NO: 174]

5' A(G or A)N(T or C)(T or C)NT(C or T)(A or G)TAN(G or C)(A or G)NANN(C or T)(C or T) 3' [SEQ ID NO: 175]

10 5' (G or A)(G or A)N(A or T)T(A or C or T)(T or A)(G or C)NTA(T or C)(G or A)AN(A or G)(A or C or G)N(T or C)T 3' [SEQ ID NO: 176]

Based on MOTIF 6:

15 5' GTNTT(T or C)(T or C)TN(T or A)(G or C)NTT(T or C)(A or C)G(A or G)GG 3' [SEQ ID NO: 177]

Based on MOTIF 7:

5' CCNAT(A or C or T)TT(T or C)TA(T or C)(G or A)(T or A)(G or T or C)GTNGA(T or C)CC 3' [SEQ ID NO: 178]

20 Based on MOTIF 8:

5' GTNGGNAT(A or C or T)GA(T or C)(G or A)(A or C)NCA 3' [SEQ ID NO: 179]

Based on MOTIF 9:

25 5' (G or A)AA(G or A)CANGC(A or G or T)AT(G or A)TCNA(G or A)(G or A)AA 3' [SEQ ID NO: 180]

5' TT(T or C)(T or C)TNGA(T or C)AT(A or C or T)GCNTG(T or C)TT 3' [SEQ ID NO: 181]

Based on MOTIF 10:

30 5' CCCAT(G or A)TC(T or C)(T or C)(T or G)NA(T or G or A)N(T or A)(G or A)(G or A)TC(A or G)TGCAT 3' [SEQ ID NO: 182]

5' ATGCA(T or C)GA(T or C)(T or C)(T or A)N(A or C or T)TN(A or C)(A or G)(A or G)GA(T or C)ATGGG 3' [SEQ ID NO: 183]

35 Based on MOTIF 11:

- 46 -

5' NA(G or A)N(G or C)(A or T)(T or C)T(T or C)NA(A or G)(C or T)TT 3' [SEQ ID NO: 184]

5' (A or T)(G or C)NAA(A or G)(T or C)TN(A or G)A(A or G)(A or T)(G or C)N(T or C)T 3' [SEQ ID NO: 185]

Based on MOTIF 12:

5' (A or G or T)(A or T)(A or T)(C or T)TCNA(G or A)N(G or C)(A or T)N(T or C)(G or T)NA(G or A) NCC 3' [SEQ ID NO: 186]

10 5' GGN(T or C)TN(A or C)(G or A)N(A or T)(G or L)N(T or C)TNGA 3' [SEQ ID NO: 187]

Once a clone encoding a candidate RPS family gene is identified, it is then determined whether such gene is capable of conferring disease-resistance to a plant host using the methods described herein or other methods well known in the art of molecular plant pathology.

A Biolistic Transient Expression Assay For Identification of Plant Resistance Genes

We have developed a functional transient expression system capable of providing a rapid and broadly applicable method for identifying and characterizing virtually any gene for its ability to confer disease-resistance to a plant cell. In brief, the assay system involves delivering by biolistic transformation a candidate plant disease-resistance gene to a plant tissue sample (e.g., a piece of tissue from a leaf) and then evaluating the expression of the gene within the tissue by appraising the presence or absence of a disease-resistance response (e.g., the hypersensitive response). This assay provides a method for identifying disease-resistance genes from a wide variety of plant species, including ones that are not amenable to genetic or transgenic studies.



- 47 -

The principle of the assay is depicted in the top portion of Figure 9. In general, plant cells carrying a mutation in the resistance gene of interest are utilized. Prior to biolistic transformation, the plant tissue is

5 infiltrated with a phytopathogenic bacterium carrying the corresponding avirulence gene. In addition, a gene to be assayed for its resistance gene activity is co-introduced by biolistics with a reporter gene. The expression of the cobombarded reporter gene serves as an indicator for

10 viability of the transformed cells. Both genes are expressed under the control of a strong and constitutive promoter. If the gene to be assayed does not complement the resistance gene function, the plant cells do not undergo a hypersensitive response (HR) and, therefore,

15 survive (Fig. 9, top panel, right). In this case, cells accumulate a large amount of the reporter gene product. If, on the other hand, a resistance gene is introduced, the plant cells recognize the signal from the avirulence-gene-carrying bacterium and undergo the HR because the

20 expressed resistance gene product complements the function (Fig. 9, top panel, left). In this case, the plant cells do not have enough time to accumulate a large amount of reporter gene product before their death. Given the transformation efficiency estimated by a proper

25 control (such as the uninfected half of the leaf), measuring the accumulation of reporter gene product can thus indicate whether the gene to be assayed complements the resistance gene function.

In one working example, we now demonstrate the

30 effectiveness of the transient expression assay, using the bacterial avirulence gene *avrRpt2* and the corresponding *Arabidopsis thaliana* resistance gene *RPS2* (Fig. 9, bottom panel). In brief, *rps2* mutant leaves, preinfected with *P. syringae* carrying *avrRpt2*, were co-

35 bombarded with two plasmids, one of which contained the

- 48 -

*RPS2* gene and the other the *Escherichia coli uidA* gene encoding  $\beta$ -glucuronidase (GUS; Jefferson et al., 1986, supra). Both the *RPS2* and *uidA* genes are located downstream of the strong constitutive 35S promoter from cauliflower mosaic virus (Odell et al., infra). If the 35S-*RPS2* construct complements the *rps2* mutation, the transformed cells rapidly undergo programmed cell death in response to the *P. syringae* carrying *avrRpt2*, and relatively little GUS activity accumulates. If the *rps2* mutation is not complemented, cell death does not occur and high levels of GUS activity accumulate. These differences in GUS activity are detected histochemically. Because the cDNA library used to identify *RPS2* was constructed in the expression vector pKEx4tr, the 35S-*RPS2* cDNA construct in pKEx4tr could be used directly in the transient assay. As shown in Fig. 11, pKEx4tr is a cDNA expression vector designed for the unidirectional insertion of cDNA inserts. Inserted cDNA is expressed under the control of the 35S cauliflower mosaic virus promoter.

Our results are shown in Fig. 9, lower panel. In this experiment, we infected one side of a leaf of an *rps2* mutant plant with *P. syringae* pv. *phaselolicola* 3121 carrying *avrRpt2* (*Psp* 3121/*avrRpt2*). *Psp* 3121 is a weak pathogen of *A. thaliana* and *Psp* 3121/*avrRpt2* can elicit an HR in a plant carrying the resistance gene *RPS2* (e.g., a wild type plant). Leaves of 5-week-old *Arabidopsis* plants were infiltrated with an appropriate bacterial suspension at a dose of  $2 \times 10^8$ /ml by hand infiltration as described (Dong et al., supra). After an incubation period (typically 2-4 hours), the leaves were bombarded using a Bio-Rad PDS-1000/He apparatus (1100 psi) after 2-4 hr of infection. Gold particles were prepared according to the instructions of the manufacturer. For each bombardment, 1.4  $\mu$ g of pKEx4tr-G, 0.1  $\mu$ g of a

- 49 -

plasmid to be tested, and 0.5 mg of 1  $\mu$ m gold particles were used. After the bombardment, the leaves were incubated in a humidity chamber at 22°C for 1 day and then subjected to a histochemical GUS staining using 5-bromo-4-chloro-3-indiyl glucuronidase (X-Gluc) at 37°C for 12 hr (Jefferson, 1987, *supra*). This staining method with X-gluc stains cells expressing GUS enzyme with a blue color. The uninfected side of the leaf serves as a control for transformation efficiency of the leaf because in a single leaf, transformation efficiency (i.e., density of transformed cells) is similar on both sides of the leaf. If transformed cells on the infected side are rapidly killed, staining of the cells on the infected side is weaker than staining on the uninfected side. When the resistance gene *RPS2* was co-introduced, the transformed cells on the infected side of the leaf showed much weaker staining than ones on the uninfected side (Fig. 10). In contrast, when an unrelated gene was co-introduced, the transformed cells on the infected side showed similar staining intensity to ones on the uninfected side (Fig. 10).

Thus, as summarized in the Table 2, *35S-RPS4* (cDNA 4), but not cDNA-5 or cDNA-6, complemented the HR phenotype of *rps2-101C*. (See Figure 1).

25

Table 2

<u>Gene Tested</u>	<u>Response (Decreased GUSActivity)<sup>a</sup></u>
$\Delta$ GUS (35S- <i>uidA</i> containing internal <i>uidA</i> deletion)	-
30 cDNA-5 (35S-AB11)	
cDNA-4 (35S-RPS2)	+
cDNA-6 (35S-CK1)	

- 50 -

<sup>a</sup>When decreased GUS activity was observed on the infiltrated side of the leaf, the response was scored as plus (Fig. 10).

Both *RPS2* cDNA-4 clones 4 and 11, corresponding to the  
5 two *RPS2* different transcript sizes, complemented the  
*rps2* mutant phenotype, indicating that both transcripts  
encode a functional product. Moreover, 35S-*RPS2* also  
complemented mutants *rps2-102C*, *rps2-101N*, and *rps2-201C*,  
further confirming that the *rps2-101C*, *rps2-102C*, *rps2-101C*  
10 *201C* and *rps2-101N* mutations are all allelic. In short,  
the cloned *RPS2* gene complemented the *rps2* mutation in  
this transient expression assay, and complementation by  
*RPS2* was observed in all four available *rps2* mutant  
stains.

15       Next we used the transient assay system to test  
the specificity of the cloned *RPS2* gene for an *avrRpt2*-  
generated signal (i.e., the "gene-for-gene" specificity  
of a *P. syringae* avirulence gene and a corresponding *A.*  
*thaliana* resistance gene (*avrRpm1* and *RPM1*,  
20 respectively)). This experiment involved the use of an  
*rps2-101 rpm1* double mutant that cannot mount an HR when  
challenged with *P. syringae* carrying *avrRpt2* or the  
unrelated avirulence gene *avrRpm1* (Debener et al., Plant  
Journal 1:289-302, 1991). As summarized in Table 3,  
25 complementation of the *rps2* mutant phenotype by 35S-*RPS2*  
was only observed in the presence of a signal generated  
by *avrRpt2*, indicating that *RPS2* does not simply  
sensitize the plant resistance response in a nonspecific  
manner.

- 51 -

**Table 3**

<u>avr Gene</u>	<u>Construct Cobombarded with 35S-uidA</u>	<u>Response<sup>a</sup></u>
None (vector only)	ΔGUS <sup>b</sup>	-
5 <i>avrRpt2</i>	ΔGUS	-
<i>avrRpm1</i>	ΔGUS	-
None (vector only)	35S-RPS2	-
<i>avrRpt2</i>	35S-RPS2	+
<i>avrRpm1</i>	35S-RPS2	-

10           <sup>a</sup>When decreased GUS activity was observed on the infiltrated side of the leaf, the response was scored as plus. (Figure 10, panel B)

<sup>b</sup>ΔGUS is 35S-uidA containing an internal deletion in the uidA gene.

15 Also as shown in Table 3, the RPS2 gene complemented the mutant phenotype when leaves were infected with Psp 3121/*avrRpt2* but not with Psp 3121/*avrRpm1*. Therefore, the RPS2 gene complemented only the *rps2* mutation; it did not the *rpm1* mutation.

20           We have also discovered that overexpression of an *rps* gene family member, e.g., *rps2* but not other genes, in the transient assay leads to apparent cell death, obviating the need to know the corresponding avirulence gene for a putative resistance gene that has been cloned.

25           Using this assay, any plant disease-resistance gene may be identified from a cDNA expression library.

In one particular example, a cDNA library is constructed in an expression vector and then introduced as described herein into a plant cultivar or its corresponding mutant  
30 plant lacking the resistance gene of interest.

Preferably, the cDNA library is divided into small pools, and each pool co-introduced with a reporter gene. If a pool contains a resistance gene clone (i.e., the pool

- 52 -

"complements" the resistance gene function), the positive pool is divided into smaller pools and the same procedure is repeated until identification of a single positive clone is ultimately achieved. This approach facilitates the cloning of any resistance gene of interest without genetic crosses or the creation of transgenics.

We now describe the cloning of another member of the RPS gene family, the Prf gene of tomato.

The initial step for the cloning of the Prf gene came from classical genetic analysis which showed that Prf was tightly linked to the tomato Pto gene (Salmeron et al., The Plant Cell 6:511-520, 1994). This prompted construction of a cosmid contig of 200 kb in length which encompassed the Pto locus. DNA probes from this contig were used to screen a tomato cDNA library constructed using tomato leaf tissue that had been infected with Pst expressing the avrPto avirulence gene as source material. Two classes of cDNAs were identified based on cross-hybridization of clones to each other. While one class corresponded to members of the Pto gene family, the other class displayed no hybridization to Pto family members. Taking the assumption (based on the aforementioned genetic analysis) that Prf might reside extremely close to the Pto gene, cDNAs from the second class were analyzed further as candidate Prf clones. These clones were hybridized to filters containing DNAs from six independent prf mutant lines that had been isolated by diepoxybutane or fast neutron treatment. In one of the fast neutron mutants, the cDNA probe revealed a 1.1 kb deletion in the genomic DNA, suggesting that the cDNA clone might in fact represent Prf. Wild-type DNA corresponding to the deletion was cloned from Prf/Prf tomato. A 5 kb region was sequenced and found to potentially encode a protein containing P-loop and leucine-rich repeat motifs, supporting the hypothesis

- 53 -

that this DNA encoded Prf. The corresponding DNA was cloned and sequenced from the fast neutron mutant plant. Sequencing this DNA confirmed the mutation to be a simple 1.1 kb deletion excising DNA between the potential P-loop and leucine-rich repeat coding regions. The gene is expressed based on RT-PCR analysis which has shown that an mRNA is transcribed from this region. The identity of the cloned DNA as the Prf gene is based on both the existence of the deletion mutation and the predicted protein sequence, which reveals patches of strong similarity to other cloned disease resistance gene products throughout the amino-terminal half (as described herein). A partial sequence of the Prf gene is shown in Figure 12.

#### 15 RPS Expression in Transgenic Plant Cells and Plants

The expression of the RPS2 genes in plants susceptible to pathogens carrying *avrRpt2* is achieved by introducing into a plant a DNA sequence containing the RPS2 gene for expression of the Rps2 polypeptide. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants are available to the public; such vectors are described in, e.g., Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985, Supp. 1987); Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; and Gelvin et al., *Plant Molecular Biology Manual*, Kluwer Academic Publishers, 1990. Typically, plant expression vectors include (1) one or more cloned plant genes under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally- or developmentally-regulated, or cell-

- 54 -

or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

5           An example of a useful plant promoter which could be used to express a plant resistance gene according to the invention is a caulimovirus promoter, e.g., the cauliflower mosaic virus (CaMV) 35S promoter. These promoters confer high levels of expression in most plant  
10 tissues, and the activity of these promoters is not dependent on virtually encoded proteins. CaMV is a source for both the 35S and 19S promoters. In most tissues of transgenic plants, the CaMV 35S promoter is a strong promoter (see, e.g., Odel et al., Nature 313:810,  
15 (1985)). The CaMV promoter is also highly active in monocots (see, e.g., Dekeyser et al., Plant Cell 2:591, (1990); Terada and Shimamoto, Mol. Gen. Genet. 220:389, (1990)).

Other useful plant promoters include, without  
20 limitation, the nonpaline synthase promoter (An et al., Plant Physiol. 88:547, (1988)) and the octopine synthase promoter (Fromm et al., Plant Cell 1:977, (1989)).

For certain applications, it may be desirable to produce the *RPS2* gene product or the *avrRpt2* gene product  
25 in an appropriate tissue, at an appropriate level, or at an appropriate developmental time. Thus, there are a variety of gene promoters, each with its own distinct characteristics embodied in its regulatory sequences, shown to be regulated in response to the environment,  
30 hormones, and/or developmental cues. These include gene promoters that are responsible for (1) heat-regulated gene expression (see, e.g., Callis et al., Plant Physiol. 88: 965, (1988)), (2) light-regulated gene expression (e.g., the pea *rbcS*-3A described by Kuhlemeier et al.,  
35 Plant Cell 1: 471, (1989); the maize *rbcS* promoter



- 55 -

- described by Schaffner and Sheen, Plant Cell 3: 997, (1991); or the chlorophyll a/b-binding protein gene found in pea described by Simpson et al., EMBO J. 4: 2723, (1985)), (3) hormone-regulated gene expression (e.g., the
- 5 abscisic acid responsive sequences from the *Em* gene of wheat described Marcotte et al., Plant Cell 1:969, (1989)), (4) wound-induced gene expression (e.g., of *wun1* described by Siebertz et al., Plant Cell 1: 961, (1989)), or (5) organ-specific gene expression (e.g., of the
- 10 tuber-specific storage protein gene described by Roshal et al., EMBO J. 6:1155, (1987); the 23-kDa zein gene from maize described by Schernthaner et al., EMBO J. 7: 1249, (1988); or the French bean  $\beta$ -phaseolin gene described by Bustos et al., Plant Cell 1:839, (1989)).
- 15 Plant expression vectors may also optionally include RNA processing signals, e.g, introns, which have been shown to be important for efficient RNA synthesis and accumulation (Callis et al., Genes and Dev. 1: 1183, (1987)). The location of the RNA splice sequences can
- 20 influence the level of transgene expression in plants. In view of this fact, an intron may be positioned upstream or downstream of an Rps2 polypeptide-encoding sequence in the transgene to modulate levels of gene expression.
- 25 In addition to the aforementioned 5' regulatory control sequences, the expression vectors may also include regulatory control regions which are generally present in the 3' regions of plant genes (Thornburg et al., Proc. Natl Acad. Sci USA 84: 744, (1987); An et
- 30 al., Plant Cell 1: 115, (1989)). For example, the 3' terminator region may be included in the expression vector to increase stability of the mRNA. One such terminator region may be derived from the PI-II terminator region of potato. In addition, other commonly

- 56 -

used terminators are derived from the octopine or nopaline synthase signals.

The plant expression vector also typically contains a dominant selectable marker gene used to  
5 identify the cells that have become transformed. Useful selectable marker genes for plant systems include genes encoding antibiotic resistance genes, for example, those encoding resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin or spectinomycin. Genes required for  
10 photosynthesis may also be used as selectable markers in photosynthetic-deficient strains. Finally, genes encoding herbicide resistance may be used as selectable markers; useful herbicide resistance genes include the bar gene encoding the enzyme phosphinothricin  
15 acetyltransferase, which confers resistance to the broad spectrum herbicide Basta® (Hoechst AG, Frankfurt, Germany).

Efficient use of selectable markers is facilitated by a determination of the susceptibility of a plant cell  
20 to a particular selectable agent and a determination of the concentration of this agent which effectively kills most, if not all, of the transformed cells. Some useful concentrations of antibiotics for tobacco transformation include, e.g., 75-100 µg/ml (kanamycin), 20-50 µg/ml  
25 (hygromycin), or 5-10 µg/ml (bleomycin). A useful strategy for selection of transformants for herbicide resistance is described, e.g., in Vasil I.K., *Cell Culture and Somatic Cell Genetics of Plants*, Vol I, II, III Laboratory Procedures and Their Applications Academic  
30 Press, New York, 1984.

It should be readily apparent to one skilled in the field of plant molecular biology that the level of gene expression is dependent not only on the combination of promoters, RNA processing signals and terminator

- 57 -

elements, but also on how these elements are used to increase the levels of gene expression.

The above exemplary techniques may be used for the expression of any gene in the RPS family.

## 5 Plant Transformation

- Upon construction of the plant expression vector, several standard methods are known for introduction of the recombinant genetic material into the host plant for the generation of a transgenic plant. These methods
- 10 include (1) *Agrobacterium*-mediated transformation (*A. tumefaciens* or *A. rhizogenes*) (see, e.g., Lichtenstein and Fuller In: *Genetic Engineering*, vol 6, PWJ Rigby, ed, London, Academic Press, 1987; and Lichtenstein, C.P., and Draper, J., In: *DNA Cloning*, Vol II, D.M. Glover, ed,
- 15 Oxford, IRI Press, 1985), (2) the particle delivery system (see, e.g., Gordon-Kamm et al., *Plant Cell* 2:603, (1990); or BioRad Technical Bulletin 1687, *supra*), (3) microinjection protocols (see, e.g., Green et al., *Plant Tissue and Cell Culture*, Academic Press, New York, 1987),
- 20 (4) polyethylene glycol (PEG) procedures (see, e.g., Draper et al., *Plant Cell Physiol* 23:451, (1982); or e.g., Zhang and Wu, *Theor. Appl. Genet.* 76:835, (1988)), (5) liposome-mediated DNA uptake (see, e.g., Freeman et al., *Plant Cell Physiol* 25: 1353, (1984)), (6)
- 25 electroporation protocols (see, e.g., Gelvin et al *supra*; Dekeyser et al. *supra*; or Fromm et al *Nature* 319: 791, (1986)), and (7) the vortexing method (see, e.g., Kindle, K., *Proc. Natl. Acad. Sci., USA* 87:1228, (1990)).

The following is an example outlining an

- 30 *Agrobacterium*-mediated plant transformation. The general process for manipulating genes to be transferred into the genome of plant cells is carried out in two phases. First, all the cloning and DNA modification steps are done in *E. coli*, and the plasmid containing the gene

- 58 -

construct of interest is transferred by conjugation into *Agrobacterium*. Second, the resulting *Agrobacterium* strain is used to transform plant cells. Thus, for the generalized plant expression vector, the plasmid contains  
5 an origin of replication that allows it to replicate in *Agrobacterium* and a high copy number origin of replication functional in *E. coli*. This permits facile production and testing of transgenes in *E. coli* prior to transfer to *Agrobacterium* for subsequent introduction  
10 into plants. Resistance genes can be carried on the vector, one for selection in bacteria, e.g., streptomycin, and the other that will express in plants, e.g., a gene encoding for kanamycin resistance or an herbicide resistance gene. Also present are restriction  
15 endonuclease sites for the addition of one or more transgenes operably linked to appropriate regulatory sequences and directional T-DNA border sequences which, when recognized by the transfer functions of *Agrobacterium*, delimit the region that will be  
20 transferred to the plant.

In another example, plant cells may be transformed by shooting into the cell tungsten microprojectiles on which cloned DNA is precipitated. In the Biolistic Apparatus (Bio-Rad, Hercules, CA) used for the shooting,  
25 a gunpowder charge (22 caliber Power Piston Tool Charge) or an air-driven blast drives a plastic macroprojectile through a gun barrel. An aliquot of a suspension of tungsten particles on which DNA has been precipitated is placed on the front of the plastic macroprojectile. The  
30 latter is fired at an acrylic stopping plate that has a hole through it that is too small for the macroprojectile to go through. As a result, the plastic macroprojectile smashes against the stopping plate and the tungsten microprojectiles continue toward their target through the  
35 hole in the plate. For the instant invention the target

- 59 -

can be any plant cell, tissue, seed, or embryo. The DNA introduced into the cell on the microprojectiles becomes integrated into either the nucleus or the chloroplast.

Transfer and expression of transgenes in plant cells is now routine practice to those skilled in the art. It has become a major tool to carry out gene expression studies and to attempt to obtain improved plant varieties of agricultural or commercial interest.

#### 10 Transgenic Plant Regeneration

Plant cells transformed with a plant expression vector can be regenerated, e.g., from single cells, callus tissue or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues and organs from almost any plant can be successfully cultured to regenerate an entire plant; such techniques are described, e.g., in Vasil supra; Green et al., supra; Weissbach and Weissbach, supra; and Gelvin et al., supra.

20 In one possible example, a vector carrying a selectable marker gene (e.g., kanamycin resistance), a cloned *RPS2* gene under the control of its own promoter and terminator or, if desired, under the control of exogenous regulatory sequences such as the 35S CaMV promoter and the nopaline synthase terminator is transformed into *Agrobacterium*. Transformation of leaf tissue with vector-containing *Agrobacterium* is carried out as described by Horsch et al. (Science 227: 1229, (1985)). Putative transformants are selected after a few weeks (e.g., 3 to 5 weeks) on plant tissue culture media containing kanamycin (e.g. 100 µg/ml). Kanamycin-resistant shoots are then placed on plant tissue culture media without hormones for root initiation. Kanamycin-resistant plants are then selected for greenhouse growth.

- 60 -

If desired, seeds from self-fertilized transgenic plants can then be sowed in a soil-less media and grown in a greenhouse. Kanamycin-resistant progeny are selected by sowing surfaced sterilized seeds on hormone-free  
5 kanamycin-containing media. Analysis for the integration of the transgene is accomplished by standard techniques (see, e.g., Ausubel et al. supra; Gelvin et al. supra).

Transgenic plants expressing the selectable marker are then screened for transmission of the transgene DNA  
10 by standard immunoblot and DNA and RNA detection techniques. Each positive transgenic plant and its transgenic progeny are unique in comparison to other transgenic plants established with the same transgene. Integration of the transgene DNA into the plant genomic  
15 DNA is in most cases random and the site of integration can profoundly effect the levels, and the tissue and developmental patterns of transgene expression. Consequently, a number of transgenic lines are usually screened for each transgene to identify and select plants  
20 with the most appropriate expression profiles.

Transgenic lines are evaluated for levels of transgene expression. Expression at the RNA level is determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA  
25 analysis are employed and include PCR amplification assays using oligonucleotide primers designed to amplify only transgene RNA templates and solution hybridization assays using transgene-specific probes (see, e.g., Ausubel et al., supra). The RNA-positive plants are then  
30 analyzed for protein expression by Western immunoblot analysis using Rps2 polypeptide-specific antibodies (see, e.g., Ausubel et al., supra). In addition, *in situ* hybridization and immunocytochemistry according to standard protocols can be done using transgene-specific

- 61 -

nucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue.

Once the Rps2 polypeptide has been expressed in any cell or in a transgenic plant (e.g., as described above), it can be isolated using any standard technique, e.g., affinity chromatography. In one example, an anti-Rps2 antibody (e.g., produced as described in Ausubel et al., supra, or by any standard technique) may be attached to a column and used to isolate the polypeptide. Lysis and fractionation of Rps2-producing cells prior to affinity chromatography may be performed by standard methods (see, e.g., Ausubel et al., supra). Once isolated, the recombinant polypeptide can, if desired, be further purified, e.g., by high performance liquid chromatography (see, e.g., Fisher, *Laboratory Techniques In Biochemistry And Molecular Biology*, Work and Burdon, eds., Elsevier, 1980).

These general techniques of polypeptide expression and purification can also be used to produce and isolate useful Rps2 fragments or analogs.

#### Antibody Production

Using a polypeptide described above (e.g., the recombinant protein or a chemically synthesized RPS peptide based on its deduced amino acid sequence), polyclonal antibodies which bind specifically to an RPS polypeptide may be produced by standard techniques (see, e.g., Ausubel et al., supra) and isolated, e.g., following peptide antigen affinity chromatography. Monoclonal antibodies can also be prepared using standard hybridoma technology (see, e.g., Kohler et al., *Nature* 256: 495, 1975; Kohler et al., *Eur. J. Immunol.* 6: 292, 1976; Hammerling et al., in *Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, N.Y., 1981; and Ausubel et al., supra).

- 62 -

Once produced, polyclonal or monoclonal antibodies are tested for specific RSP polypeptide recognition by Western blot or immunoprecipitation analysis (by methods described in Ausubel et al., supra). Antibodies which specifically recognize a RPS polypeptide are considered to be useful in the invention; such antibodies may be used, e.g., for screening recombinant expression libraries as described in Ausubel et al., supra. Exemplary peptides (derived from Rps2) for antibody production include:

LKFSYDNLESDDL [SEQ ID NO: 188]  
GVYGPGGVGKTTLMQS [SEQ ID NO: 189]  
GGLPLALITLGGAM [SEQ ID NO: 190]

#### Use

Introduction of RPS2 into a transformed plant cell provides for resistance to bacterial pathogens carrying the *avrRpt2* avirulence gene. For example, transgenic plants of the instant invention expressing RPS2 might be used to alter, simply and inexpensively, the disease resistance of plants normally susceptible to plant pathogens carrying the avirulence gene, *avrRpt2*.

The invention also provides for broad-spectrum pathogen resistance by mimicking the natural mechanism of host resistance. First, the RPS2 transgene is expressed in plant cells at a sufficiently high level to initiate the plant defense response constitutively in the absence of signals from the pathogen. The level of expression associated with plant defense response initiation is determined by measuring the levels of defense response gene expression as described in Dong et al., supra. Second, the RPS2 transgene is expressed by a controllable promoter such as a tissue-specific promoter, cell-type specific promoter or by a promoter that is induced by an external signal or agent thus limiting the temporal and



- 63 -

tissue expression of a defense response. Finally, the *RPS2* gene product is co-expressed with the *avrRpt2* gene product. The *RPS2* gene is expressed by its natural promoter, by a constitutively expressed promoter such as  
5 the CaMV 35S promoter, by a tissue-specific or cell-type specific promoter, or by a promoter that is activated by an external signal or agent. Co-expression of *RPS2* and *avrRpt2* will mimic the production of gene products associated with the initiation of the plant defense  
10 response and provide resistance to pathogens in the absence of specific resistance gene-avirulence gene corresponding pairs in the host plant and pathogen.

The invention also provides for expression in plant cells of a nucleic acid having the sequence of Fig.  
15 2 or the expression of a degenerate variant thereof encoding the amino acid sequence of open reading frame "a" of Fig. 2.

The invention further provides for the isolation of nucleic acid sequences having about 50% or greater  
20 sequence identity to *RPS2* by using the *RPS2* sequence of Fig. 2 or a portion thereof greater than 9 nucleic acids in length, and preferably greater than about 18 nucleic acids in length as a probe. Appropriate reduced hybridization stringency conditions are utilized to  
25 isolate DNA sequences having about 50% or greater sequence identity to the *RPS2* sequence of Fig. 2.

Also provided by the invention are short conserved regions characteristic of *RPS* disease resistance genes. These conserved regions provide oligonucleotide sequences  
30 useful for the production of hybridization probes and PCR primers for the isolation of other plant disease-resistance genes.

Both the *RPS2* gene and related *RPS* family genes provide disease resistance to plants, especially crop  
35 plants, most especially important crop plants such as

- 64 -

tomato, pepper, maize, wheat, rice and legumes such as soybean and bean, or any plant which is susceptible to pathogens carrying an avirulence gene, e.g., the *avrRpt2* avirulence gene. Such pathogens include, but are not  
5 limited to, *Pseudomonas syringae* strains.

The invention also includes any biologically active fragment or analog of an Rps2 polypeptide. By "biologically active" is meant possessing any in vivo activity which is characteristic of the Rps2 polypeptide  
10 shown in Fig. 2. A useful Rps2 fragment or Rps2 analog is one which exhibits a biological activity in any biological assay for disease resistance gene product activity, for example, those assays described by Dong et al. (1991), supra; Yu et al. (1993) supra; Kunkel et al.  
15 (1993) supra; and Whalen et al. (1991). In particular, a biologically active Rps2 polypeptide fragment or analog is capable of providing substantial resistance to plant pathogens carrying the *avrRpt2* avirulence gene. By substantial resistance is meant at least partial  
20 reduction in susceptibility to plant pathogens carrying the *avrRpt2* gene.

Preferred analogs include Rps2 polypeptides (or biologically active fragments thereof) whose sequences differ from the wild-type sequence only by conservative  
25 amino acid substitutions, for example, substitution of one amino acid for another with similar characteristics (e.g., valine for glycine, arginine for lysine, etc.) or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not abolish the  
30 polypeptide's biological activity.

Analogous can differ from naturally occurring Rps2 polypeptide in amino acid sequence or can be modified in ways that do not involve sequence, or both. Analogous of the invention will generally exhibit at least 70%,  
35 preferably 80%, more preferably 90%, and most preferably

- 65 -

95% or even 99%, homology with a segment of 20 amino acid residues, preferably 40 amino acid residues, or more preferably the entire sequence of a naturally occurring Rps2 polypeptide sequence.

- 5           Alterations in primary sequence include genetic variants, both natural and induced. Also included are analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g.,  $\beta$  or  
10  $\gamma$  amino acids. Also included in the invention are Rps2 polypeptides modified by *in vivo* chemical derivatization of polypeptides, including acetylation, methylation, phosphorylation, carboxylation, or glycosylation.

- In addition to substantially full-length  
15 polypeptides, the invention also includes biologically active fragments of the polypeptides. As used herein, the term "fragment", as applied to a polypeptide, will ordinarily be at least 20 residues, more typically at least 40 residues, and preferably at least 60 residues in  
20 length. Fragments of Rps2 polypeptide can be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a biological activity of Rps2 can be assessed by those methods described herein. Also included in the invention are  
25 Rps2 polypeptides containing residues that are not required for biological activity of the peptide, e.g., those added by alternative mRNA splicing or alternative protein processing events.

Other embodiments are within the following claims.

- 66 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Ausubel, Frederick M.  
Staskawicz, Brian J.  
Brent, Andrew F.  
Dahlbeck, Douglas  
Katagiri, Fumiaki  
Kunkel, Barbara N.  
Mindrinis, Michael N.  
Yu, Guo-Liang

(ii) TITLE OF INVENTION: RPS2 GENE AND USES THEREOF

(iii) NUMBER OF SEQUENCES: 201

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Fish & Richardson  
(B) STREET: 225 Franklin Street Suite 3100  
(C) CITY: Boston  
(D) STATE: MA  
(E) COUNTRY: USA  
(F) ZIP: 02110-2904

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30B

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/227,360  
(B) FILING DATE: 13-APR-1994  
(C) CLASSIFICATION:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Clark, Paul T.  
(B) REGISTRATION NUMBER: 30,162  
(C) REFERENCE/DOCKET NUMBER: 00786/230001

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (617) 542-5070  
(B) TELEFAX: (617) 542-8906  
(C) TELEX: 100254

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2903 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AAGTAAAGA AAGAGCGAGA AATCATCGAA ATGGATTCA TCTCATCTCT TATCGTTGGC	60
TGTGCTCAGG TGTGTGTGA ATCTATGAAT ATGGCGGAGA GAAGAGGACA TAAGACTGAT	120

- 67 -

CTTAGACAAG CCATCACTGA TCTTGAAACA GCCATCGGTG ACTTGAAGGC CATACGTGAT	180
GACCTGACTT TACGGATCCA ACAAGACGGT CTAGAGGGAC GAAGCTGCTC AAATCGTGCC	240
AGAGAGTGGC TTAGTGCGGT GCAAGTAACG GAGACTAAAA CAGCCCTACT TTTAGTGAGG	300
TTTAGGCGTC GGGAAACAGAG GACGCGAATG AGGAGGAGAT ACCTCAGTTG TTTCCGTTGT	360
GGCGACTACA AACTGTGCAA GAAGGTTTCT GCCATATTGA AGAGCATTGG TGAGCTGAGA	420
GAACGCTCTG AAGCTATCAA AACAGATGGC GGGTCAATTC AAGTAACTTG TAGAGAGATA	480
CCCATCAAGT CCGTTGTCGG AAATACCACG ATGATGGAAC AGGTTTTGGA ATTTCTCAGT	540
GAAGAAGAAG AAAGAGGAAT CATTGGTGTT TATGGACCTG GTGGGGTTGG GAAGACAACG	600
TTAATGCAGA GCATTAACAA CGAGCTGATC ACAAAGGAC ATCAGTATGA TGTACTGATT	660
TGGGTTCAAA TGTCCAGAGA ATTCGGCGAG TGTACAATTC AGCAAGCCGT TGGAGCACGG	720
TTGGGTTTAT CTTGGGACGA GAAGGAGACC GGCGAAAACA GAGCTTTGAA GATATACAGA	780
GCTTTGAGAC AGAAACGTTT CTTGTTGTTG CTAGATGATG TCTGGGAAGA GATAGACTTG	840
GAGAAAAGT GAGTTCCTCG ACCTGACAGG GAAAACAAAT GCAAGGTGAT GTTCACGACA	900
CGGTCTATAG CATTATGCAA CAATATGGGT GCGGAATACA AGTTGAGAGT GGAGTTTCTG	960
GAGAAGAAAC ACGCGTGGGA GCTGTTCTGT AGTAAGGTAT GGAGAAAAGA TCTTTTAGAG	1020
TCATCATCAA TTCGCCGGCT CGCGGAGATT ATAGTGAGTA AATGTGGAGG ATTGCCACTA	1080
GCGTTGATCA CTTTAGGAGG AGCCATGGCT CATAGAGAGA CAGAAGAAGA GTGGATCCAT	1140
GCTAGTGAAG TTCTGACTAG ATTTCCAGCA GAGATGAAGG GTATGAACTA TGTATTTGCC	1200
CTTTTGAAAT TCAGCTACGA CAACCTCGAG AGTGATCTGC TTCGGTCTTG TTTCTTGAC	1260
TGCGCTTTAT TCCCAGAAGA ACATTCTATA GAGATCGAGC AGCTTGTTGA GTACTGGGTC	1320
GGCGAAGGGT TTCTCACCAG CTCCCATGGC GTTAACACCA TTTACAAGGG ATATTTTCTC	1380
ATTGGGGATC TGAAAGCGGC ATGTTTGTTG GAAACCGGAG ATGAGAAAAC ACAGGTGAAG	1440
ATGCATAATG TGGTCAGAAG CTTTGCAATG TGGATGGCAT CTGAACAGGG GACTTATAAG	1500
GAGCTGATCC TAGTTGAGCC TAGCATGGGA CATACTGAAG CTCCTAAAGC AGAAAAGTGG	1560
CGACAAGCGT TGGTGATCTC ATTGTTAGAT AACAGAATCC AGACCTTGCC TGAAAACTC	1620
ATATGCCCCG AACTGACAAC ACTGATGCTC CAACAGAACA GCTCTTTGAA GAAGATTCCA	1680
ACAGGGTTTT TCATGCATAT GCCTGTTCTC AGAGTCTTGG ACTTGTCGTT CACAAGTATC	1740
ACTGAGATTG CGTTGTCTAT CAAGTATTG GTGGAGTTGT ATCATCTGTC TATGTCAGGA	1800
ACAAAGATAA GTGTATTGCC ACAGGAGCTT GGAATCTTA GAAAAGTGA GCATCTGGAC	1860
CTACAAAGAA CTCAGTTTCT TCAGACGATC CCACGAGATG CCATATGTTG GCTGAGCAAG	1920
CTCAGGTTT TGAAGTTGTA CTACAGTTAC GCCGGTTGGG AACTGCAGAG CTTTGGAGAA	1980
GATGAAGCAG AAGAACTCGG ATTCGCTGAC TTGGAATACT TGGAAAACCT AACCACACTC	2040

- 68 -

```

GGTATCACTG TTCTCTCATT GGAGACCCTA AAAACTCTCT TCGAGTTCGG TGCTTTGCAT 2100
AAACATATAC AGCATCTCCA CGTTGAAGAG TGCAATGAAC TCCTCTACTT CAATCTCCCA 2160
TCACTCACTA ACCATGGCAG GAACCTGAGA AGACTTAGCA TTAAGAGTTG CCATGACTTG 2220
GAGTACCTGG TCACACCCGC AGATTTTGAA AATGATTGGC TTCCGAGTCT AGAGGTTCTG 2280
ACGTTACACA GCCTTCACAA CTTAACCAGA GTGTGGGGAA ATTCTGTAAG CCAAGATTGT 2340
CTGCGGAATA TCCGTTGCAT AAACATTTCA CACTGCAACA AGCTGAAGAA TGTCTCATGG 2400
GTTTCAGAAAC TCCCAAAGCT AGAGGTGATT GAACTGTTCC ACTGCAGAGA GATAGAGGAA 2460
TTGATAAGCG AACACGAGAG TCCATCCGTC GAAGATCCAA CATTGTTCCC AAGCCTGAAG 2520
ACCTTGAGAA CTAGGGATCT GCCAGAACTA AACAGCATCC TCCCATCTCG ATTTTCATTC 2580
CAAAAAGTTG AAACATTAGT CATCACAAAT TGCCCCAGAG TTAAGAAACT GCCGTTTCAG 2640
GAGAGGAGGA CCCAGATGAA CTTGCCAACA GTTTATTGTG AGGAGAAATG GTGGAAAGCA 2700
CTGGAAAAAG ATCAACCAA CGAAGAGCTT TGTTATTTAC CGCGCTTTGT TCCAAATTGA 2760
TATAAGAGCT AAGAGCACTC TGTACAAATA TGTCCATTCA TAAGTAGCAG GAAGCCAGGA 2820
AGGTTGTTCC AGTGAAGTCA TCAACTTTCC ACATAGCCAC AAAACTAGAG ATTATGTAAT 2880
CATAAAAACC AAATATCCG CGA 2903

```

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 885 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Lys Lys Glu Arg Glu Ile Ile Glu Met Asp Phe Ile Ser Ser Leu Ile
1           5           10           15
Val Gly Cys Ala Gln Val Leu Cys Glu Ser Met Asn Met Ala Glu Arg
20           25           30
Arg Gly His Lys Thr Asp Leu Arg Gln Ala Ile Thr Asp Leu Arg Ile
35           40           45
Gln Gln Asp Gly Leu Glu Gly Arg Ser Cys Ser Asn Arg Ala Arg Glu
50           55           60
Trp Leu Ser Ala Val Gln Val Thr Glu Thr Lys Thr Ala Leu Leu Leu
65           70           75           80
Val Arg Phe Arg Arg Arg Glu Gln Arg Thr Arg Met Arg Arg Arg Tyr
85           90           95

```

- 69 -

Leu Ser Cys Phe Gly Cys Ala Asp Tyr Lys Leu Cys Lys Lys Val Ser  
 100 105 110  
 Ala Ile Leu Lys Ser Ile Gly Glu Leu Arg Glu Arg Ser Glu Ala Ile  
 115 120 125  
 Lys Thr Asp Gly Gly Ser Ile Gln Val Thr Cys Arg Glu Ile Pro Ile  
 130 135 140  
 Lys Ser Val Val Gly Asn Thr Thr Met Met Glu Gln Val Leu Glu Phe  
 145 150 155 160  
 Leu Ser Glu Glu Glu Glu Arg Gly Ile Ile Gly Val Tyr Gly Pro Gly  
 165 170 175  
 Gly Val Gly Lys Thr Thr Leu Met Gln Ser Ile Asn Asn Glu Leu Ile  
 180 185 190  
 Thr Lys Gly His Gln Tyr Asp Val Leu Ile Trp Val Gln Met Ser Arg  
 195 200 205  
 Glu Phe Gly Glu Cys Thr Ile Gln Gln Ala Val Gly Ala Arg Leu Gly  
 210 215 220  
 Leu Ser Trp Asp Glu Lys Glu Thr Gly Glu Asn Arg Ala Leu Lys Ile  
 225 230 235 240  
 Tyr Arg Ala Leu Arg Gln Lys Arg Phe Leu Leu Leu Leu Asp Asp Val  
 245 250 255  
 Trp Glu Glu Ile Asp Leu Glu Lys Thr Gly Val Pro Arg Pro Asp Arg  
 260 265 270  
 Glu Asn Lys Cys Lys Val Met Phe Thr Thr Arg Ser Ile Ala Leu Cys  
 275 280 285  
 Asn Asn Met Gly Ala Glu Tyr Lys Leu Arg Val Glu Phe Leu Glu Lys  
 290 295 300  
 Lys His Ala Trp Glu Leu Phe Cys Ser Lys Val Trp Arg Lys Asp Leu  
 305 310 315 320  
 Leu Glu Ser Ser Ser Ile Arg Arg Leu Ala Glu Ile Ile Val Ser Lys  
 325 330 335  
 Cys Gly Gly Leu Pro Leu Ala Leu Ile Thr Leu Gly Gly Ala Met Ala  
 340 345 350  
 His Arg Glu Thr Glu Glu Glu Trp Ile His Ala Ser Glu Val Leu Thr  
 355 360 365  
 Arg Phe Pro Ala Glu Met Lys Gly Met Asn Tyr Val Phe Ala Leu Leu  
 370 375 380  
 Lys Phe Ser Tyr Asp Asn Leu Glu Ser Asp Leu Leu Arg Ser Cys Phe  
 385 390 395 400  
 Leu Tyr Cys Ala Leu Phe Pro Glu Glu His Ser Ile Glu Ile Glu Gln  
 405 410 415  
 Leu Val Glu Tyr Trp Val Gly Glu Gly Phe Leu Thr Ser Ser His Gly  
 420 425 430

- 70 -

Val Asn Thr Ile Tyr Lys Gly Tyr Phe Leu Ile Gly Asp Leu Lys Ala  
 435 440 445  
 Ala Cys Leu Leu Glu Thr Gly Asp Glu Lys Thr Gln Val Lys Met His  
 450 455 460  
 Asn Val Val Arg Ser Phe Ala Leu Trp Met Ala Ser Glu Gln Gly Thr  
 465 470 475 480  
 Tyr Lys Glu Leu Ile Leu Val Glu Pro Ser Met Gly His Thr Glu Ala  
 485 490 495  
 Pro Lys Ala Glu Asn Trp Arg Gln Ala Leu Val Ile Ser Leu Leu Asp  
 500 505 510  
 Asn Arg Ile Gln Thr Leu Pro Glu Lys Leu Ile Cys Pro Lys Leu Thr  
 515 520 525  
 Thr Leu Met Leu Gln Gln Asn Ser Ser Leu Lys Lys Ile Pro Thr Gly  
 530 535 540  
 Phe Phe Met His Met Pro Val Leu Arg Val Leu Asp Leu Ser Phe Thr  
 545 550 555 560  
 Ser Ile Thr Glu Ile Pro Leu Ser Ile Lys Tyr Leu Val Glu Leu Tyr  
 565 570 575  
 His Leu Ser Met Ser Gly Thr Lys Ile Ser Val Leu Pro Gln Glu Leu  
 580 585 590  
 Gly Asn Leu Arg Lys Leu Lys His Leu Asp Leu Gln Arg Thr Gln Phe  
 595 600 605  
 Leu Gln Thr Ile Pro Arg Asp Ala Ile Cys Trp Leu Ser Lys Leu Glu  
 610 615 620  
 Val Leu Asn Leu Tyr Tyr Ser Tyr Ala Gly Trp Glu Leu Gln Ser Phe  
 625 630 635 640  
 Gly Glu Asp Glu Ala Glu Glu Leu Gly Phe Ala Asp Leu Glu Tyr Leu  
 645 650 655  
 Glu Asn Leu Thr Thr Leu Gly Ile Thr Val Leu Ser Leu Glu Thr Leu  
 660 665 670  
 Lys Thr Leu Phe Glu Phe Gly Ala Leu His Lys His Ile Gln His Leu  
 675 680 685  
 His Val Glu Glu Cys Asn Glu Leu Leu Tyr Phe Asn Leu Pro Ser Leu  
 690 695 700  
 Thr Asn His Gly Arg Asn Leu Arg Arg Leu Ser Ile Lys Ser Cys His  
 705 710 715 720  
 Asp Leu Glu Tyr Leu Val Thr Pro Ala Asp Phe Glu Asn Asp Trp Leu  
 725 730 735  
 Pro Ser Leu Glu Val Leu Thr Leu His Ser Leu His Asn Leu Arg Cys  
 740 745 750  
 Ile Asn Ile Ser His Cys Asn Lys Leu Lys Asn Val Ser Trp Val Gln  
 755 760 765



- 71 -

Lys Leu Pro Lys Leu Glu Val Ile Glu Leu Phe Asp Cys Arg Glu Ile  
 770 775 780  
 Glu Glu Leu Ile Ser Glu His Glu Ser Pro Ser Val Glu Asp Pro Thr  
 785 790 795 800  
 Leu Phe Pro Ser Leu Lys Thr Leu Arg Thr Arg Asp Leu Pro Glu Leu  
 805 810 815  
 Asn Ser Ile Leu Pro Ser Arg Phe Ser Phe Gln Lys Val Glu Thr Leu  
 820 825 830  
 Val Ile Thr Asn Cys Pro Arg Val Lys Lys Leu Pro Phe Gln Glu Arg  
 835 840 845  
 Arg Thr Gln Met Asn Leu Pro Thr Val Tyr Cys Glu Glu Lys Trp Trp  
 850 855 860  
 Lys Ala Leu Glu Lys Asp Gln Pro Asn Glu Glu Leu Cys Tyr Leu Pro  
 865 870 875 880  
 Arg Phe Val Pro Asn  
 885

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Glu His Ser Val Gln Ile Cys Pro Phe Ile Ser Ser Arg Lys Pro Gly  
 1 5 10 15  
 Arg Leu Phe Gln  
 20

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Ser His Gln Leu Ser Thr  
 1 5

- 72 -

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Arg	Leu	Cys	Asn	His	Lys	Asn	Gln	Thr	Ile	Arg
1					5				10	

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ser	Lys	Arg	Lys	Ser	Glu	Lys	Ser	Ser	Lys	Trp	Ile	Ser	Ser	His	Leu
1				5					10					15	
Leu	Ser	Leu	Ala	Val	Leu	Arg	Cys	Cys	Val	Asn	Leu				
			20					25							

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Ile	Trp	Arg	Arg	Glu	Glu	Asp	Ile	Arg	Leu	Ile	Leu	Asp	Lys	Pro	Ser
1				5					10					15	
Leu	Ile	Leu	Lys	Gln	Pro	Ser	Val	Thr							
			20					25							

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:

- 73 -

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Arg Pro Tyr Val Met Thr  
1 5

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Leu Tyr Gly Ser Asn Lys Thr Val  
1 5

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Arg Asp Glu Ala Ala Gln Ile Val Pro Glu Ser Gly Leu Val Arg Cys  
1 5 10 15  
Lys

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 74 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Arg Arg Leu Lys Gln Pro Tyr Phe  
1 5

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Gly Leu Gly Val Gly Asn Arg Gly Arg Glu  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Gly Gly Asp Thr Ser Val Val Ser Val Val Pro Thr Thr Asn Cys Ala  
1 5 10 15

Arg Arg Phe Leu Pro Tyr  
20

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Arg Ala Leu Val Ser

- 75 -

1

5

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Glu	Asn	Ala	Leu	Lys	Leu	Ser	Lys	Gln	Met	Ala	Gly	Gln	Phe	Lys
1				5					10					15

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Leu	Val	Glu	Arg	Tyr	Pro	Ser	Ser	Pro	Leu	Ser	Glu	Ile	Pro	Arg
1				5					10					15

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Trp	Asn	Arg	Phe	Trp	Asn	Phe	Ser	Val	Lys	Lys	Lys	Lys	Glu	Glu	Ser
1				5					10					15	

Leu	Val	Phe	Met	Asp	Leu	Val	Gly	Leu	Gly	Arg	Gln	Arg
			20					25				

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids

- 76 -

- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Cys Arg Ala Leu Thr Thr Ser  
1 5

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ser Gln Lys Asp Ile Ser Met Met Tyr  
1 5

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Phe Gly Phe Lys Cys Pro Glu Asn Ser Ala Ser Val Gln Phe Ser Lys  
1 5 10 15  
Pro Leu Glu His Gly Trp Val Tyr Leu Gly Thr Arg Arg Arg Pro Ala  
20 25 30  
Lys Thr Glu Leu  
35

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

- 77 -

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Arg Tyr Thr Glu Leu  
1 5

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Asp Arg Asn Val Ser Cys Cys Cys  
1 5

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met Met Ser Gly Lys Arg  
1 5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Thr Trp Arg Lys Leu Glu Phe Leu Asp Leu Thr Gly Lys Thr Asn Ala  
1 5 10 15

- 78 -

Arg

## (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 6 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Cys Ser Arg His Gly Leu  
 1 5

## (2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

His Tyr Ala Thr Ile Trp Val Arg Asn Thr Ser  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Glu Trp Ser Phe Trp Arg Arg Asn Thr Arg Gly Ser Cys Ser Val Val  
 1 5 10 15

Arg Tyr Gly Glu Lys Ile Phe  
 20

## (2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:



- 79 -

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Ser His His Gln Phe Ala Gly Ser Arg Arg Leu  
1 5 10

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Val Asn Val Glu Asp Cys His  
1 5

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Glu Glu Pro Trp Leu Ile Glu Arg Gln Lys Lys Ser Gly Ser Met Leu  
1 5 10 15

Val Lys Phe

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 6 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 80 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Leu Asp Phe Gln Gln Arg  
 1 5

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Thr Met Tyr Leu Pro Phe  
 1 5

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 27 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Asn Ser Ala Thr Thr Thr Ser Arg Val Ile Cys Phe Gly Leu Val Ser  
 1 5 10 15

Cys Thr Ala Leu Tyr Ser Gln Lys Asn Ile Leu  
 20 25

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Arg Ser Ser Ser Leu Leu Ser Thr Gly Ser Ala Lys Gly Phe Ser Pro

- 81 -

1                      5                      10                      15  
 Ala Pro Met Ala Leu Thr Pro Phe Thr Arg Asp Ile Phe Ser Leu Gly  
                     20                      25                      30  
 Ile

## (2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Lys Arg His Val Cys Trp Lys Pro Glu Met Arg Lys His Arg  
 1                      5                      10

## (2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 22 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Arg Cys Ile Met Trp Ser Glu Ala Leu His Cys Gly Trp His Leu Asn  
 1                      5                      10                      15

Arg Gly Leu Ile Arg Ser  
 20

## (2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Leu Ser Leu Ala Trp Asp Ile Leu Lys Leu Leu Lys Gln Lys Thr Gly  
 1                      5                      10                      15

- 82 -

Asp Lys Arg Trp  
20

## (2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Ile Thr Glu Ser Arg Pro Cys Leu Lys Asn Ser Tyr Ala Arg Asn  
 1 5 10 15

## (2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Cys Ser Asn Arg Thr Ala Leu  
 1 5

## (2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 46 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Arg Arg Phe Gln Gln Gly Phe Ser Cys Ile Cys Leu Phe Ser Glu Ser  
 1 5 10 15

Trp Thr Cys Arg Ser Gln Val Ser Leu Arg Phe Arg Cys Leu Ser Ser  
 20 25 30

Ile Trp Trp Ser Cys Ile Ile Cys Leu Cys Gln Glu Gln Arg  
 35 40 45

- 83 -

## (2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Val	Tyr	Cys	His	Arg	Ser	Leu	Gly	Ile	Leu	Glu	Asn
1				5					10		

## (2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Ser	Ile	Trp	Thr	Tyr	Lys	Glu	Leu	Ser	Phe	Phe	Arg	Arg	Ser	His	Glu
1				5					10					15	
Met	Pro	Tyr	Val	Gly											
			20												

## (2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Ala	Ser	Ser	Arg	Phe
1			5	

## (2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant

- 84 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Thr	Cys	Thr	Thr	Val	Thr	Pro	Val	Gly	Asn	Cys	Arg	Ala	Leu	Glu	Lys
1				5				10					15		
Met	Lys	Gln	Lys	Asn	Ser	Asp	Ser	Leu	Thr	Trp	Asn	Thr	Trp	Lys	Thr
		20						25					30		

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Pro	His	Ser	Val	Ser	Leu	Phe	Ser	His	Trp	Arg	Pro
1				5					10		

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Lys	Leu	Ser	Ser	Ser	Ser	Val	Leu	Cys	Ile	Asn	Ile	Tyr	Ser	Ile	Ser
1					5				10					15	
Thr	Leu	Lys	Ser	Ala	Met	Asn	Ser	Ser	Thr	Ser	Ile	Ser	His	His	Ser
		20						25					30		
Leu	Thr	Met	Ala	Gly	Thr										
		35													

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid

- 85 -

- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Glu Asp Leu Ala Leu Lys Val Ala Met Thr Trp Ser Thr Trp Ser His  
1 5 10 15

Pro Gln Ile Leu Lys Met Ile Gly Phe Arg Val  
20 25

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Arg Tyr Thr Ala Phe Thr Thr  
1 5

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Pro Glu Cys Gly Glu Ile Leu  
1 5

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 86 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Ala Lys Ile Val Cys Gly Ile Ser Val Ala  
1 5 10

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Thr Phe His Thr Ala Thr Ser  
1 5

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Phe Arg Asn Ser Gln Ser  
1 5

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Leu Asn Cys Ser Thr Ala Glu Arg  
1 5

(2) INFORMATION FOR SEQ ID NO:54:



- 87 -

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 16 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Ala Asn Thr Arg Val His Pro Ser Lys Ile Gln His Cys Ser Gln Ala  
1                      5                      10                      15

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Glu Leu Gly Ile Cys Gln Asn  
1                      5

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Thr Ala Ser Ser His Leu Asp Phe His Ser Lys Lys Leu Lys His  
1                      5                      10                      15

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 88 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Ser Ser Gln Ile Ala Pro Glu Leu Arg Asn Cys Arg Phe Arg Arg Gly  
 1 5 10 15  
 Gly Pro Arg

## (2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 47 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Thr Cys Gln Gln Phe Ile Val Arg Arg Asn Gly Gly Lys His Trp Lys  
 1 5 10 15  
 Lys Ile Asn Gln Thr Lys Ser Phe Val Ile Tyr Arg Ala Leu Phe Gln  
 20 25 30  
 Ile Asp Ile Arg Ala Lys Ser Thr Leu Tyr Lys Tyr Val His Ser  
 35 40 45

## (2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Val Ala Gly Ser Gln Glu Gly Cys Ser Ser Glu Val Ile Asn Phe Pro  
 1 5 10 15  
 His Ser His Lys Thr Arg Asp Tyr Val Ile Ile Lys Thr Lys Leu Ser  
 20 25 30  
 Ala

## (2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 25 amino acids  
 (B) TYPE: amino acid

- 89 -

- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Val	Lys	Glu	Arg	Ala	Arg	Asn	His	Arg	Asn	Gly	Phe	His	Leu	Ile	Ser
1				5				10					15		
Tyr	Arg	Trp	Leu	Cys	Ser	Gly	Val	Val							
			20				25								

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ile	Tyr	Glu	Tyr	Gly	Gly	Glu	Lys	Arg	Thr
1				5					10

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Leu	Glu	Gly	His	Thr
1			5	

(2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 90 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Pro Asp Phe Thr Asp Pro Thr Arg Arg Ser Arg Gly Thr Lys Leu Leu  
1 5 10 15  
Lys Ser Cys Gln Arg Val Ala  
20

## (2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Cys Gly Ala Ser Asn Gly Asp  
1 5

## (2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Asn Ser Pro Thr Phe Ser Glu Val  
1 5

## (2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 35 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Ala Ser Gly Thr Glu Asp Ala Asn Glu Glu Glu Ile Pro Gln Leu Phe  
1 5 10 15

- 91 -

Arg Leu Cys Arg Leu Gln Thr Val Gln Glu Gly Phe Cys His Ile Glu  
                   20                  25                  30  
 Glu His Trp  
                   35

## (2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Ala Glu Arg Thr Leu  
 1                  5

## (2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Ser Tyr Gln Asn Arg Trp Arg Val Asn Ser Ser Asn Leu  
 1                  5                  10

## (2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 22 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Arg Asp Thr His Gln Val Arg Cys Arg Lys Tyr His Asp Asp Gly Thr  
 1                  5                  10                  15  
 Gly Phe Gly Ile Ser Gln  
                   20

- 92 -

## (2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

```

Arg Arg Arg Lys Arg Asn His Trp Cys Leu Trp Thr Trp Trp Gly Trp
1           5           10           15

Glu Asp Asn Val Asn Ala Glu His
                20

```

## (2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

```

Gln Arg Ala Asp His Lys Arg Thr Ser Val
1           5           10

```

## (2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 55 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

```

Cys Thr Asp Leu Gly Ser Asn Val Gln Arg Ile Arg Arg Val Tyr Asn
1           5           10           15

Ser Ala Ser Arg Trp Ser Thr Val Gly Phe Ile Leu Gly Arg Glu Gly
                20           25           30

Asp Arg Arg Lys Gln Ser Phe Glu Asp Ile Gln Ser Phe Glu Thr Glu
35           40           45

```

- 93 -

Thr Phe Leu Val Val Ala Arg  
50 55

## (2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Cys Leu Gly Arg Asp Arg Leu Gly Glu Asn Trp Ser Ser Ser Thr  
1 5 10 15

## (2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Arg Asp Arg Arg Arg Val Asp Pro Cys  
1 5

## (2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 41 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Gln Gly Lys Gln Met Gln Gly Asp Val His Asp Thr Val Tyr Ser Ile  
1 5 10 15

Met Gln Gln Tyr Gly Cys Gly Ile Gln Val Glu Ser Gly Val Ser Gly  
20 25 30

Glu Glu Thr Arg Val Gly Ala Val Leu  
35 40

- 94 -

## (2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Gly Met Glu Lys Arg Ser Phe Arg Val Ile Ile Asn Ser Pro Ala Arg  
1 5 10 15  
Gly Asp Tyr Ser Glu  
20

## (2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Met Trp Arg Ile Ala Thr Ser Val Asp His Phe Arg Arg Ser His Gly  
1 5 10 15  
Ser

## (2) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Ile Ser Ser Arg Asp Glu Gly Tyr Glu Leu Cys Ile Cys Pro Phe Glu  
1 5 10 15  
Ile Gln Leu Arg Gln Pro Arg Glu  
20



- 95 -

## (2) INFORMATION FOR SEQ ID NO:79:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Ser Ala Ser Val Leu Phe Leu Val Leu Arg Phe Ile Pro Arg Arg Thr  
1 5 10 15  
Phe Tyr Arg Asp Arg Ala Ala Cys  
20

## (2) INFORMATION FOR SEQ ID NO:80:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Val Leu Gly Arg Arg Arg Val Ser His Gln Leu Pro Trp Arg  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:81:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

His His Leu Gln Gly Ile Phe Ser His Trp Gly Ser Glu Ser Gly Met  
1 5 10 15  
Phe Val Gly Asn Arg Arg  
20

## (2) INFORMATION FOR SEQ ID NO:82:

- (i) SEQUENCE CHARACTERISTICS:

- 96 -

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Glu Asn Thr Gly Glu Asp Ala  
1 5

(2) INFORMATION FOR SEQ ID NO:83:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 43 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Lys Thr His Met Pro Glu Thr Asp Asn Thr Asp Ala Pro Thr Glu Gly  
1 5 10 15

Leu Phe Glu Glu Asp Ser Asn Arg Val Phe His Ala Tyr Ala Cys Ser  
20 25 30

Gln Ser Leu Gly Leu Val Val His Lys Tyr His  
35 40

(2) INFORMATION FOR SEQ ID NO:84:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Cys Gly Gln Lys Leu Cys Ile Val Asp Gly Ile  
1 5 10

(2) INFORMATION FOR SEQ ID NO:85:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant

- 97 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Gly Ala Asp Pro Ser  
1 5

(2) INFORMATION FOR SEQ ID NO:86:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Ser Arg Lys Leu Ala Thr Ser Val Gly Asp Leu Ile Val Arg  
1 5 10

(2) INFORMATION FOR SEQ ID NO:87:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Gln Asn Pro Asp Leu Ala  
1 5

(2) INFORMATION FOR SEQ ID NO:88:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

- 98 -

Asp	Ser	Val	Val	Tyr	Gln	Val	Phe	Gly	Gly	Val	Val	Ser	Ser	Val	Tyr
1				5					10					15	
Val	Arg	Asn	Lys	Asp	Lys	Cys	Ile	Ala	Thr	Gly	Ala	Trp	Glu	Ser	
			20					25					30		

## (2) INFORMATION FOR SEQ ID NO:89:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 47 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Lys	Thr	Glu	Ala	Ser	Gly	Pro	Thr	Lys	Asn	Ser	Val	Ser	Ser	Asp	Asp
1				5				10						15	
Pro	Thr	Arg	Cys	His	Met	Leu	Ala	Glu	Gln	Ala	Arg	Gly	Ser	Glu	Leu
			20					25					30		
Val	Leu	Gln	Leu	Arg	Arg	Leu	Gly	Thr	Ala	Glu	Leu	Trp	Arg	Arg	
			35				40					45			

## (2) INFORMATION FOR SEQ ID NO:90:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Ser	Arg	Arg	Thr	Arg	Ile	Arg
1				5		

## (2) INFORMATION FOR SEQ ID NO:91:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

- 99 -

Leu Gly Ile Leu Gly Lys Pro Asn His Thr Arg Tyr His Cys Ser Leu  
 1 5 10 15

Ile Gly Asp Pro Lys Asn Ser Leu Arg Val Arg Cys Phe Ala  
 20 25 30

## (2) INFORMATION FOR SEQ ID NO:92:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Thr Tyr Thr Ala Ser Pro Arg  
 1 5

## (2) INFORMATION FOR SEQ ID NO:93:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

Thr Pro Leu Leu Gln Ser Pro Ile Thr His  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:94:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Pro Trp Gln Glu Pro Glu Lys Thr  
 1 5

## (2) INFORMATION FOR SEQ ID NO:95:

- (i) SEQUENCE CHARACTERISTICS:

- 100 -

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Leu Gly Val Pro Gly His Thr Arg Arg Phe  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 58 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Leu Ala Ser Glu Ser Arg Gly Ser Asp Val Thr Gln Pro Ser Gln Leu  
 1 5 10 15  
 Asn Gln Ser Val Gly Lys Phe Cys Lys Pro Arg Leu Ser Ala Glu Tyr  
 20 25 30  
 Pro Leu His Lys His Phe Thr Leu Gln Gln Ala Glu Glu Cys Leu Met  
 35 40 45  
 Gly Ser Glu Thr Pro Lys Ala Arg Gly Asp  
 50 55

(2) INFORMATION FOR SEQ ID NO:97:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Thr Val Arg Leu Gln Arg Asp Arg Gly Ile Asp Lys Arg Thr Arg Glu  
 1 5 10 15  
 Ser Ile Arg Arg Arg Ser Asn Ile Val Pro Lys Pro Glu Asp Leu Glu  
 20 25 30

- 101 -

Asn

## (2) INFORMATION FOR SEQ ID NO:98:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Gly	Ser	Ala	Arg	Thr	Lys	Gln	His	Pro	Pro	Ile	Ser	Ile	Phe	Ile	Pro
1				5				10					15		

Lys Ser

## (2) INFORMATION FOR SEQ ID NO:99:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

Asn	Ile	Ser	His	His	Lys	Leu	Pro	Gln	Ser
1				5				10	

## (2) INFORMATION FOR SEQ ID NO:100:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Glu	Thr	Ala	Val	Ser	Gly	Glu	Glu	Asp	Pro	Asp	Glu	Leu	Ala	Asn	Ser
1				5				10					15		

Leu Leu

- 102 -

## (2) INFORMATION FOR SEQ ID NO:101:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

Thr Ser His His  
1

## (2) INFORMATION FOR SEQ ID NO:102:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

Glu Leu Arg Ala Leu Cys Thr Asn Met Ser Ile His Lys  
1                      5                      10

## (2) INFORMATION FOR SEQ ID NO:103:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

Gln Glu Ala Arg Lys Val Val Pro Val Lys Ser Ser Thr Phe His Ile  
1                      5                      10                      15  
Ala Thr Lys Leu Glu Ile Met  
20

## (2) INFORMATION FOR SEQ ID NO:104:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 6 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant



- 103 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Lys Pro Asn Tyr Pro Arg  
 1 5

(2) INFORMATION FOR SEQ ID NO:105:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1491 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

ATCGATTGAT CTCTGGCTCA GTGCGAGTAG TCCATTTGAG AGCAGTCGTA GCCCCGCGTG	60
GCGCATCATG GAGCTATTTG GAATTTTCGC AGGGTTATCG ATTCGTAGTG GGAACCCATT	120
CATTGTTTGG AACCACCAAC GGACGACTTA ACAAGCTCCC CGAGGTGCAT GATGAAAATT	180
GCTCCAGTTG CCATAAATCA CAGCCCGCTC AGCAGGGAGG TCCCCTCACA CGCGGCACCC	240
ACTCAGGCAA AGCAAACCAA CCTTCAATCT GAAGCTGGCG ATTTAGATGC AAGAAAAAGT	300
AGCGCTTCAA GCCCGGAAAC CCGCGCATTG CTCGCTACTA AGACAGTACT CGGGAGACAC	360
AAGATAGAGG TTCCGGCCTT TGGAGGGTGG TTCAAAAAGA AATCATCTAA GCACGAGACG	420
GGCGGTTCAA GTGCCAACGC AGATAGTTCTG AGCGTGGCTT CCGATTCCAC CGAAAAACCT	480
TTGTTCCGTC TCACGCACGT TCCTTACGTA TCCCAAGGTA ATGAGCGAAT GGGATGTTGG	540
TATGCCTGCG CAAGAATGGT TGGCCATTCT GTCGAAGCTG GGCCTCGCCT AGGGCTGCCG	600
GAGCTCTATG AGGGAAGGGA GCGCCAGCT GGGCTACAAG ATTTTTCAGA TGTAGAAAGG	660
TTTATTCACA ATGAAGGATT AACTCGGGTA GACCTTCCAG ACAATGAGAG ATTTACACAC	720
GAAGAGTTGG GTGCACTGTT GTATAAGCAC GGGCCGATTA TATTTGGGTG GAAAACTCCG	780
AATGACAGCT GGCACATGTC GGTCTCACT GGTGTGATA AAGAGACGTC GTCCATTACT	840
TTTCACGATC CCCGACAGGG GCCGGACCTA GCAATGCCGC TCGATTACTT TAATCAGCGA	900
TTGGCATGGC AGGTTCCACA CGCAATGCTC TACCGCTAAG TAGCAGGGTA TCTTCACGTG	960
GCGGCATCAT GACAAGCCCA TGATGCCGCC AGCAGCTACC TGAATGCCGT CTGGCTTTTT	1020
GGTCCCTATT GTCGTATCCG GAAGATGACG TCAAAGAATC TCGGCAAGAG CTTTCTTGCT	1080

- 104 -

CGACTCCTCA GCTTCCGGAT CGATCAGGTC GCTTGCCAGA GCGCGCTTGT CCATGAGCAT 1140  
 CTGCCACAGC TGCTGGTCGA TGGTGTCTC AGCTAAAGGG ATTTTGACGA CAACCATGCG 1200  
 CAACTGCCCG TTGCGATACG CTCGATCCTG AAGCCCCGGT GTCCATGGCA GCCCCAAGAA 1260  
 AAAGACATAG TTCGCCGCTG TGAGGTTGTA GCCTGTGCCG GCGGCCGACC TGGTCCCGAT 1320  
 AAACACCCTG CAGTCCGGAT CCTGCTGGAA AGCATCAATC GCCTTCTGCC GCTTCTTGGG 1380  
 CGAGTCACTG CCCACCAACG TCACGCACCC GACGCCAAGC TTGAGGCAGT GCTCCCGCAA 1440  
 CGTGGCCACG GATTCTGAT ACTCGCAGAA GAGGATCACC TTGTCGTCGA C 1491

## (2) INFORMATION FOR SEQ ID NO:106:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 255 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

Met Lys Ile Ala Pro Val Ala Ile Asn His Ser Pro Leu Ser Arg Glu  
 1 5 10 15  
 Val Pro Ser His Ala Ala Pro Thr Gln Ala Lys Gln Thr Asn Leu Gln  
 20 25 30  
 Ser Glu Ala Gly Asp Leu Asp Ala Arg Lys Ser Ser Ala Ser Ser Pro  
 35 40 45  
 Glu Thr Arg Ala Leu Leu Ala Thr Lys Thr Val Leu Gly Arg His Lys  
 50 55 60  
 Ile Glu Val Pro Ala Phe Gly Gly Trp Phe Lys Lys Lys Ser Ser Lys  
 65 70 75 80  
 His Glu Thr Gly Gly Ser Ser Ala Asn Ala Asp Ser Ser Ser Val Ala  
 85 90 95  
 Ser Asp Ser Thr Glu Lys Pro Leu Phe Arg Leu Thr His Val Pro Tyr  
 100 105 110  
 Val Ser Gln Gly Asn Glu Arg Met Gly Cys Trp Tyr Ala Cys Ala Arg  
 115 120 125  
 Met Val Gly His Ser Val Glu Ala Gly Pro Arg Leu Gly Leu Pro Glu  
 130 135 140  
 Leu Tyr Glu Gly Arg Glu Ala Pro Ala Gly Leu Gln Asp Phe Ser Asp  
 145 150 155 160  
 Val Glu Arg Phe Ile His Asn Glu Gly Leu Thr Arg Val Asp Leu Pro  
 165 170 175  
 Asp Asn Glu Arg Phe Thr His Glu Glu Leu Gly Ala Leu Leu Tyr Lys

- 105 -

180	185	190
His Gly Pro Ile Ile Phe Gly Trp Lys Thr Pro Asn Asp Ser Trp His		
195	200	205
Met Ser Val Leu Thr Gly Val Asp Lys Glu Thr Ser Ser Ile Thr Phe		
210	215	220
His Asp Pro Arg Gln Gly Pro Asp Leu Ala Met Pro Leu Asp Tyr Phe		
225	230	235
Asn Gln Arg Leu Ala Trp Gln Val Pro His Ala Met Leu Tyr Arg		
245	250	255

## (2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1209 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

Met Asn Pro Ser Gly Ser Phe Pro Ser Val Glu Tyr Glu Val Phe Leu		
1	5	10
Ser Phe Arg Gly Pro Asp Thr Arg Glu Gln Phe Thr Asp Phe Leu Tyr		
20	25	30
Gln Ser Leu Arg Arg Tyr Lys Ile His Thr Phe Arg Asp Asp Asp Glu		
35	40	45
Leu Leu Lys Gly Lys Glu Ile Gly Pro Asn Leu Leu Arg Ala Ile Asp		
50	55	60
Gln Ser Lys Ile Tyr Val Pro Ile Ile Ser Ser Gly Tyr Ala Asp Ser		
65	70	75
Lys Trp Cys Leu Met Glu Leu Ala Glu Ile Val Arg Arg Gln Glu Glu		
85	90	95
Asp Pro Arg Arg Ile Ile Leu Pro Ile Phe Tyr Met Val Asp Pro Ser		
100	105	110
Asp Val Arg His Gln Thr Gly Cys Tyr Lys Lys Ala Phe Arg Lys His		
115	120	125
Ala Asn Lys Phe Asp Gly Gln Thr Ile Gln Asn Trp Lys Asp Ala Leu		
130	135	140
Lys Lys Val Gly Asp Leu Lys Gly Trp His Ile Gly Lys Asn Asp Lys		
145	150	155
Gln Gly Ala Ile Ala Asp Lys Val Ser Ala Asp Ile Trp Ser His Ile		
165	170	175

- 106 -

Ser Lys Glu Asn Leu Ile Leu Glu Thr Asp Glu Leu Val Gly Ile Asp  
 180 185 190  
 Asp His Ile Thr Ala Val Leu Glu Lys Leu Ser Leu Asp Ser Glu Asn  
 195 200 205  
 Val Thr Met Val Gly Leu Tyr Gly Met Gly Gly Ile Gly Lys Thr Thr  
 210 215 220  
 Thr Ala Lys Ala Val Tyr Asn Lys Ile Ser Ser Cys Phe Asp Cys Cys  
 225 230 235 240  
 Cys Phe Ile Asp Asn Ile Arg Glu Thr Gln Glu Lys Asp Gly Val Val  
 245 250 255  
 Val Leu Gln Lys Lys Leu Val Ser Glu Ile Leu Arg Ile Asp Ser Gly  
 260 265 270  
 Ser Val Gly Phe Asn Asn Asp Ser Gly Gly Arg Lys Thr Ile Lys Glu  
 275 280 285  
 Arg Val Ser Arg Phe Lys Ile Leu Val Val Leu Asp Asp Val Asp Glu  
 290 295 300  
 Lys Phe Lys Phe Glu Asp Met Leu Gly Ser Pro Lys Asp Phe Ile Ser  
 305 310 315 320  
 Gln Ser Arg Phe Ile Ile Thr Ser Arg Ser Met Arg Val Leu Gly Thr  
 325 330 335  
 Leu Asn Glu Asn Gln Cys Lys Leu Tyr Glu Val Gly Ser Met Ser Lys  
 340 345 350  
 Pro Arg Ser Leu Glu Leu Phe Ser Lys His Ala Phe Lys Lys Asn Thr  
 355 360 365  
 Pro Pro Ser Ser Tyr Tyr Glu Thr Leu Ala Asn Asp Val Val Asp Thr  
 370 375 380  
 Thr Ala Gly Leu Pro Leu Thr Leu Lys Val Ile Gly Ser Leu Leu Phe  
 385 390 395 400  
 Lys Gln Glu Ile Ala Val Trp Glu Asp Thr Leu Glu Gln Leu Arg Arg  
 405 410 415  
 Thr Leu Asn Leu Asp Glu Val Tyr Asp Arg Leu Lys Ile Ser Tyr Asp  
 420 425 430  
 Ala Leu Asn Pro Glu Ala Lys Glu Ile Phe Leu Asp Ile Ala Cys Phe  
 435 440 445  
 Phe Ile Gly Gln Asn Lys Glu Glu Pro Tyr Tyr Met Trp Thr Asp Cys  
 450 455 460  
 Asn Phe Tyr Pro Ala Ser Asn Ile Ile Phe Leu Ile Gln Arg Cys Met  
 465 470 475 480  
 Ile Gln Val Gly Asp Asp Asp Glu Phe Lys Met His Asp Gln Leu Arg  
 485 490 495  
 Asp Met Gly Arg Glu Ile Val Arg Arg Glu Asp Val Leu Pro Trp Lys  
 500 505 510

- 107 -

Ser Arg Ile Trp Ser Ala Glu Glu Gly Ile Asp Leu Leu Leu Asn Lys  
 515 520 525  
 Arg Lys Gly Ser Ser Lys Val Lys Ala Ile Ser Ile Pro Trp Gly Val  
 530 535 540  
 Lys Tyr Glu Phe Lys Ser Glu Cys Phe Leu Asn Leu Ser Glu Leu Arg  
 545 550 555 560  
 Tyr Leu His Ala Arg Glu Ala Met Leu Thr Gly Asp Phe Asn Asn Leu  
 565 570 575  
 Leu Pro Asn Leu Lys Trp Leu Glu Leu Pro Phe Tyr Lys His Gly Glu  
 580 585 590  
 Asp Asp Pro Pro Leu Thr Asn Tyr Thr Met Lys Asn Leu Ile Ile Val  
 595 600 605  
 Ile Leu Glu His Ser His Ile Thr Ala Asp Asp Trp Gly Gly Trp Arg  
 610 615 620  
 His Met Met Lys Met Ala Glu Arg Leu Lys Val Val Arg Leu Ala Ser  
 625 630 635 640  
 Asn Tyr Ser Leu Tyr Gly Arg Arg Val Arg Leu Ser Asp Cys Trp Arg  
 645 650 655  
 Phe Pro Lys Ser Ile Glu Val Leu Ser Met Thr Ala Ile Glu Met Asp  
 660 665 670  
 Glu Val Asp Ile Gly Glu Leu Lys Lys Leu Lys Thr Leu Val Leu Lys  
 675 680 685  
 Pro Cys Pro Ile Gln Lys Ile Ser Gly Gly Thr Phe Gly Met Leu Lys  
 690 695 700  
 Gly Leu Arg Glu Leu Cys Leu Glu Phe Asn Trp Gly Thr Asn Leu Arg  
 705 710 715 720  
 Glu Val Val Ala Asp Ile Gly Gln Leu Ser Ser Leu Lys Val Leu Lys  
 725 730 735  
 Thr Gly Ala Lys Glu Val Glu Ile Asn Glu Phe Pro Leu Gly Leu Lys  
 740 745 750  
 Thr Glu Leu Ser Thr Ser Ser Arg Ile Pro Asn Asn Leu Ser Gln Leu  
 755 760 765  
 Leu Asp Leu Glu Val Leu Lys Val Tyr Asp Cys Lys Asp Gly Phe Asp  
 770 775 780  
 Met Pro Pro Ala Ser Pro Ser Glu Asp Glu Ser Ser Val Trp Trp Lys  
 785 790 795 800  
 Val Ser Lys Leu Lys Ser Leu Gln Leu Glu Lys Thr Arg Ile Asn Val  
 805 810 815  
 Asn Val Val Asp Asp Ala Ser Ser Gly Gly His Leu Pro Arg Tyr Leu  
 820 825 830  
 Leu Pro Thr Ser Leu Thr Tyr Leu Lys Ile Tyr Gln Cys Thr Glu Pro  
 835 840 845

- 108 -

Thr Trp Leu Pro Gly Ile Glu Asn Leu Glu Asn Leu Thr Ser Leu Glu  
 850 855 860  
 Val Asn Asp Ile Phe Gln Thr Leu Gly Gly Asp Leu Asp Gly Leu Gln  
 865 870 875 880  
 Gly Leu Arg Ser Leu Glu Ile Leu Arg Ile Arg Lys Val Asn Gly Leu  
 885 890 895  
 Ala Arg Ile Lys Gly Leu Lys Asp Leu Leu Cys Ser Ser Thr Cys Lys  
 900 905 910  
 Leu Arg Lys Phe Tyr Ile Thr Glu Cys Pro Asp Leu Ile Glu Leu Leu  
 915 920 925  
 Pro Cys Glu Leu Gly Val Gln Thr Val Val Val Pro Ser Met Ala Glu  
 930 935 940  
 Leu Thr Ile Arg Asp Cys Pro Arg Leu Glu Val Gly Pro Met Ile Arg  
 945 950 955 960  
 Ser Leu Pro Lys Phe Pro Met Leu Lys Lys Leu Asp Leu Ala Val Ala  
 965 970 975  
 Asn Ile Thr Lys Glu Glu Asp Leu Asp Ala Ile Gly Ser Leu Glu Glu  
 980 985 990  
 Leu Val Ser Leu Glu Leu Glu Leu Asp Asp Thr Ser Ser Gly Ile Glu  
 995 1000 1005  
 Arg Ile Val Ser Ser Ser Lys Leu Gln Lys Leu Thr Thr Leu Val Val  
 1010 1015 1020  
 Lys Val Pro Ser Leu Arg Glu Ile Glu Gly Leu Glu Glu Leu Lys Ser  
 1025 1030 1035 1040  
 Leu Gln Asp Leu Tyr Leu Glu Gly Cys Thr Ser Leu Gly Arg Leu Pro  
 1045 1050 1055  
 Leu Glu Lys Leu Lys Glu Leu Asp Ile Gly Gly Cys Pro Asp Leu Thr  
 1060 1065 1070  
 Glu Leu Val Gln Thr Val Val Ala Val Pro Ser Leu Arg Gly Leu Thr  
 1075 1080 1085  
 Ile Arg Asp Cys Pro Arg Leu Glu Val Gly Pro Met Ile Gln Ser Leu  
 1090 1095 1100  
 Pro Lys Phe Pro Met Leu Asn Glu Leu Thr Leu Ser Met Val Asn Ile  
 1105 1110 1115 1120  
 Thr Lys Glu Asp Glu Leu Glu Val Leu Gly Ser Leu Glu Glu Leu Asp  
 1125 1130 1135  
 Ser Leu Glu Leu Thr Leu Asp Asp Thr Cys Ser Ser Ile Glu Arg Ile  
 1140 1145 1150  
 Ser Phe Leu Ser Lys Leu Gln Lys Leu Thr Thr Leu Ile Val Glu Val  
 1155 1160 1165  
 Pro Ser Leu Arg Glu Ile Glu Gly Leu Ala Glu Leu Lys Ser Leu Arg  
 1170 1175 1180

- 109 -

Ile Leu Tyr Leu Glu Gly Cys Thr Ser Leu Glu Arg Leu Trp Pro Asp  
 1185 1190 1195 1200

Gln Gln Gln Leu Gly Ser Leu Lys Asn  
 1205

## (2) INFORMATION FOR SEQ ID NO:108:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1143 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

Met Ala Ser Ser Ser Ser Ser Arg Trp Ser Tyr Asp Val Phe Leu  
 1 5 10 15

Ser Phe Arg Gly Glu Asp Thr Arg Lys Thr Phe Thr Ser His Leu Tyr  
 20 25 30

Glu Val Leu Asn Asp Lys Gly Ile Lys Thr Phe Gln Asp Asp Lys Arg  
 35 40 45

Leu Glu Tyr Gly Ala Thr Ile Pro Gly Glu Leu Cys Lys Ala Ile Glu  
 50 55 60

Glu Ser Gln Phe Ala Ile Val Val Phe Ser Glu Asn Tyr Ala Thr Ser  
 65 70 75 80

Arg Trp Cys Leu Asn Glu Leu Val Lys Ile Met Glu Cys Lys Thr Arg  
 85 90 95

Phe Lys Gln Thr Val Ile Pro Ile Phe Tyr Asp Val Asp Pro Ser His  
 100 105 110

Val Arg Asn Gln Lys Glu Ser Phe Ala Lys Ala Phe Glu Glu His Glu  
 115 120 125

Thr Lys Tyr Lys Asp Asp Val Glu Gly Ile Gln Arg Trp Arg Ile Ala  
 130 135 140

Leu Asn Glu Ala Ala Asn Leu Lys Gly Ser Cys Asp Asn Arg Asp Lys  
 145 150 155 160

Thr Asp Ala Asp Cys Ile Arg Gln Ile Val Asp Gln Ile Ser Ser Lys  
 165 170 175

Leu Cys Lys Ile Ser Leu Ser Tyr Leu Gln Asn Ile Val Gly Ile Asp  
 180 185 190

Thr His Leu Glu Lys Ile Glu Ser Leu Leu Glu Ile Gly Ile Asn Gly  
 195 200 205

Val Arg Ile Met Gly Ile Trp Gly Met Gly Gly Val Gly Lys Thr Thr  
 210 215 220

Ile Ala Arg Ala Ile Phe Asp Thr Leu Leu Gly Arg Met Asp Ser Ser

- 110 -

225		230		235		240									
Tyr	Gln	Phe	Asp	Gly	Ala	Cys	Phe	Leu	Lys	Asp	Ile	Lys	Glu	Asn	Lys
				245					250					255	
Arg	Gly	Met	His	Ser	Leu	Gln	Asn	Ala	Leu	Leu	Ser	Glu	Leu	Leu	Arg
			260					265					270		
Glu	Lys	Ala	Asn	Tyr	Asn	Asn	Glu	Glu	Asp	Gly	Lys	His	Gln	Met	Ala
		275					280					285			
Ser	Arg	Leu	Arg	Ser	Lys	Lys	Val	Leu	Ile	Val	Leu	Asp	Asp	Ile	Asp
	290					295					300				
Asn	Lys	Asp	His	Tyr	Leu	Glu	Tyr	Leu	Ala	Gly	Asp	Leu	Asp	Trp	Phe
305					310					315					320
Gly	Asn	Gly	Ser	Arg	Ile	Ile	Ile	Thr	Thr	Arg	Asp	Lys	His	Leu	Ile
				325					330					335	
Glu	Lys	Asn	Asp	Ile	Ile	Tyr	Glu	Val	Thr	Ala	Leu	Pro	Asp	His	Glu
			340					345					350		
Ser	Ile	Gln	Leu	Phe	Lys	Gln	His	Ala	Phe	Gly	Lys	Glu	Val	Pro	Asn
		355					360					365			
Glu	Asn	Phe	Glu	Lys	Leu	Ser	Leu	Glu	Val	Val	Asn	Tyr	Ala	Lys	Gly
	370					375					380				
Leu	Pro	Leu	Ala	Leu	Lys	Val	Trp	Gly	Ser	Leu	Leu	His	Asn	Leu	Arg
385					390					395					400
Leu	Thr	Glu	Trp	Lys	Ser	Ala	Ile	Glu	His	Met	Lys	Asn	Asn	Ser	Tyr
				405					410					415	
Ser	Gly	Ile	Ile	Asp	Lys	Leu	Lys	Ile	Ser	Tyr	Asp	Gly	Leu	Glu	Pro
			420					425					430		
Lys	Gln	Gln	Glu	Met	Phe	Leu	Asp	Ile	Ala	Cys	Phe	Leu	Arg	Gly	Glu
		435					440					445			
Glu	Lys	Asp	Tyr	Ile	Leu	Gln	Ile	Leu	Glu	Ser	Cys	His	Ile	Gly	Ala
	450					455					460				
Glu	Tyr	Gly	Leu	Arg	Ile	Leu	Ile	Asp	Lys	Ser	Leu	Val	Phe	Ile	Ser
465					470					475					480
Glu	Tyr	Asn	Gln	Val	Gln	Met	His	Asp	Leu	Ile	Gln	Asp	Met	Gly	Lys
			485						490					495	
Tyr	Ile	Val	Asn	Phe	Gln	Lys	Asp	Pro	Gly	Glu	Arg	Ser	Arg	Leu	Trp
			500					505					510		
Leu	Ala	Lys	Glu	Val	Glu	Glu	Val	Met	Ser	Asn	Asn	Thr	Gly	Thr	Met
		515					520					525			
Ala	Met	Glu	Ala	Ile	Trp	Val	Ser	Ser	Tyr	Ser	Ser	Thr	Leu	Arg	Phe
	530					535					540				
Ser	Asn	Gln	Ala	Val	Lys	Asn	Met	Lys	Arg	Leu	Arg	Val	Phe	Asn	Met
545					550					555					560
Gly	Arg	Ser	Ser	Thr	His	Tyr	Ala	Ile	Asp	Tyr	Leu	Pro	Asn	Asn	Leu



- 111 -

				565				570				575			
Arg	Cys	Phe	Val 580	Cys	Thr	Asn	Tyr	Pro 585	Trp	Glu	Ser	Phe	Pro 590	Ser	Thr
Phe	Glu	Leu 595	Lys	Met	Leu	Val	His 600	Leu	Gln	Leu	Arg	His 605	Asn	Ser	Leu
Arg	His 610	Leu	Trp	Thr	Glu	Thr 615	Lys	His	Leu	Pro	Ser 620	Leu	Arg	Arg	Ile
Asp 625	Leu	Ser	Trp	Ser	Lys 630	Arg	Leu	Thr	Arg	Thr 635	Pro	Asp	Phe	Thr	Gly 640
Met	Pro	Asn	Leu	Glu 645	Tyr	Val	Asn	Leu	Tyr 650	Gln	Cys	Ser	Asn	Leu	Glu
Glu	Val	His	His 660	Ser	Leu	Gly	Cys	Cys 665	Ser	Lys	Val	Ile	Gly 670	Leu	Tyr
Leu	Asn	Asp 675	Cys	Lys	Ser	Leu	Lys 680	Arg	Phe	Pro	Cys	Val 685	Asn	Val	Glu
Ser	Leu	Glu 690	Tyr	Leu	Gly	Leu 695	Arg	Ser	Cys	Asp 700	Ser	Leu	Glu	Lys	Leu
Pro 705	Glu	Ile	Tyr	Gly	Arg 710	Met	Lys	Pro	Glu	Ile 715	Gln	Ile	His	Met	Gln 720
Gly	Ser	Gly	Ile	Arg 725	Glu	Leu	Pro	Ser	Ser 730	Ile	Phe	Gln	Tyr	Lys 735	Thr
His	Val	Thr	Lys 740	Leu	Leu	Leu	Trp	Asn 745	Met	Lys	Asn	Leu	Val 750	Ala	Leu
Pro	Ser	Ser 755	Ile	Cys	Arg	Leu	Lys 760	Ser	Leu	Val	Ser 765	Leu	Ser	Val	Ser
Gly	Cys 770	Ser	Lys	Leu	Glu	Ser 775	Leu	Pro	Glu	Glu	Ile 780	Gly	Asp	Leu	Asp
Asn 785	Leu	Arg	Val	Phe	Asp 790	Ala	Ser	Asp	Thr	Leu 795	Ile	Leu	Arg	Pro	Pro 800
Ser	Ser	Ile	Ile	Arg 805	Leu	Asn	Lys	Leu	Ile 810	Ile	Leu	Met	Phe	Arg 815	Gly
Phe	Lys	Asp	Gly 820	Val	His	Phe	Glu	Phe 825	Pro	Pro	Val	Ala	Glu 830	Gly	Leu
His	Ser 835	Leu	Glu	Tyr	Leu	Asn 840	Leu	Ser	Tyr	Cys	Asn 845	Leu	Ile	Asp	Gly
Gly 850	Leu	Pro	Glu	Glu	Ile	Gly 855	Ser	Leu	Ser	Ser	Leu 860	Lys	Lys	Leu	Asp
Leu 865	Ser	Arg	Asn	Asn	Phe 870	Glu	His	Leu	Pro	Ser 875	Ser	Ile	Ala	Gln	Leu 880
Gly	Ala	Leu	Gln	Ser 885	Leu	Asp	Leu	Lys	Asp 890	Cys	Gln	Arg	Leu	Thr 895	Gln
Leu	Pro	Glu	Leu	Pro	Pro	Glu	Leu	Asn	Glu	Leu	His	Val	Asp	Cys	His

- 112 -

900	905	910
Met Ala Leu Lys Phe Ile His Tyr Leu Val Thr Lys Arg Lys Lys Leu 915 920 925		
His Arg Val Lys Leu Asp Asp Ala His Asn Asp Thr Met Tyr Asn Leu 930 935 940		
Phe Ala Tyr Thr Met Phe Gln Asn Ile Ser Ser Met Arg His Asp Ile 945 950 955 960		
Ser Ala Ser Asp Ser Leu Ser Leu Thr Val Phe Thr Gly Gln Pro Tyr 965 970 975		
Pro Glu Lys Ile Pro Ser Trp Phe His His Gln Gly Trp Asp Ser Ser 980 985 990		
Val Ser Val Asn Leu Pro Glu Asn Trp Tyr Ile Pro Asp Lys Phe Leu 995 1000 1005		
Gly Phe Ala Val Cys Tyr Ser Arg Ser Leu Ile Asp Thr Thr Ala His 1010 1015 1020		
Leu Ile Pro Val Cys Asp Asp Lys Met Ser Arg Met Thr Gln Lys Leu 1025 1030 1035 1040		
Ala Leu Ser Glu Cys Asp Thr Glu Ser Ser Asn Tyr Ser Glu Trp Asp 1045 1050 1055		
Ile His Phe Phe Phe Val Pro Phe Ala Gly Leu Trp Asp Thr Ser Lys 1060 1065 1070		
Ala Asn Gly Lys Thr Pro Asn Asp Tyr Gly Ile Ile Arg Leu Ser Phe 1075 1080 1085		
Ser Gly Glu Glu Lys Met Tyr Gly Arg Leu Arg Leu Tyr Lys Glu Gly 1090 1095 1100		
Pro Glu Val Asn Ala Leu Leu Gln Met Arg Glu Asn Ser Asn Glu Pro 1105 1110 1115 1120		
Thr Glu His Ser Thr Gly Ile Arg Arg Thr Gln Tyr Asn Asn Arg Thr 1125 1130 1135		
Ser Phe Tyr Glu Leu Ile Asn 1140		

## (2) INFORMATION FOR SEQ ID NO:109:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 429 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Leu Arg Ser Lys Leu Asp Leu Ile Ile Asp Leu Lys His Gln Ile Glu
1 5 10 15

- 113 -

Ser Val Lys Glu Gly Leu Leu Cys Leu Arg Ser Phe Ile Asp His Phe  
 20 25 30  
 Ser Glu Ser Tyr Val Glu His Asp Glu Ala Cys Gly Leu Ile Ala Arg  
 35 40 45  
 Val Ser Val Met Ala Tyr Lys Ala Glu Tyr Val Ile Asp Ser Cys Leu  
 50 55 60  
 Ala Tyr Ser His Pro Leu Trp Tyr Lys Val Leu Trp Ile Ser Glu Val  
 65 70 75 80  
 Leu Glu Asn Ile Lys Leu Val Asn Lys Val Val Gly Glu Thr Cys Glu  
 85 90 95  
 Arg Arg Asn Thr Glu Val Thr Val His Glu Val Ala Lys Thr Thr Thr  
 100 105 110  
 Asn Val Ala Pro Ser Phe Ser Ala Tyr Thr Gln Arg Ala Asn Glu Glu  
 115 120 125  
 Met Glu Gly Phe Gln Asp Thr Ile Asp Glu Leu Lys Asp Lys Leu Leu  
 130 135 140  
 Gly Gly Ser Pro Glu Leu Asp Val Ile Ser Ile Val Gly Met Pro Gly  
 145 150 155 160  
 Leu Gly Lys Thr Thr Leu Ala Lys Lys Ile Tyr Asn Asp Pro Glu Val  
 165 170 175  
 Thr Ser Arg Phe Asp Val His Ala Gln Cys Val Val Thr Gln Leu Tyr  
 180 185 190  
 Ser Trp Arg Glu Leu Leu Leu Thr Ile Leu Asn Asp Val Leu Glu Pro  
 195 200 205  
 Ser Asp Arg Asn Glu Lys Glu Asp Gly Glu Ile Ala Asp Glu Leu Arg  
 210 215 220  
 Arg Phe Leu Leu Thr Lys Arg Phe Leu Ile Leu Ile Asp Asp Val Trp  
 225 230 235 240  
 Asp Tyr Lys Val Trp Asp Asn Leu Cys Met Cys Phe Ser Asp Val Ser  
 245 250 255  
 Asn Arg Ser Arg Ile Ile Leu Thr Thr Arg Leu Asn Asp Val Ala Glu  
 260 265 270  
 Tyr Val Lys Cys Glu Ser Asp Pro His His Leu Arg Leu Phe Arg Asp  
 275 280 285  
 Asp Glu Ser Trp Thr Leu Leu Gln Lys Glu Val Phe Gln Gly Glu Ser  
 290 295 300  
 Cys Pro Pro Glu Leu Glu Asp Val Gly Phe Glu Ile Ser Lys Ser Cys  
 305 310 315 320  
 Arg Gly Leu Pro Leu Ser Val Val Leu Val Ala Gly Val Leu Lys Gln  
 325 330 335  
 Lys Lys Lys Thr Leu Asp Ser Trp Lys Val Val Glu Gln Ser Leu Ser  
 340 345 350

- 114 -

Ser Gln Arg Ile Gly Ser Leu Glu Glu Ser Ile Ser Ile Ile Gly Phe  
 355 360 365

Ser Tyr Lys Asn Leu Pro His Tyr Leu Lys Pro Cys Phe Leu Tyr Phe  
 370 375 380

Gly Gly Phe Leu Gln Gly Lys Asp Ile His Asp Ser Lys Met Thr Lys  
 385 390 395 400

Leu Trp Val Ala Glu Glu Phe Val Gln Ala Asn Asn Glu Lys Gly Gln  
 405 410 415

Glu Asp Thr Arg Thr Arg Phe Leu Gly Arg Ser Tyr Trp  
 420 425

## (2) INFORMATION FOR SEQ ID NO:110:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

Gly Met Gly Gly Ile Gly Lys Thr Thr Thr Ala  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:111:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

Gly Met Gly Gly Val Gly Lys Thr Thr Ile Ala  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:112:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

- 115 -

Gly Met Pro Gly Leu Gly Lys Thr Thr Leu Ala  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:113:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

Gly Pro Gly Gly Val Gly Lys Thr Thr Leu Met  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:114:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Phe Lys Ile Leu Val Val Leu Asp Asp Val Asp  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:115:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

Lys Lys Val Leu Ile Val Leu Asp Asp Ile Asp  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:116:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 116 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Lys Arg Phe Leu Ile Leu Ile Asp Asp Val Trp  
1 5 10

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Lys Arg Phe Leu Leu Leu Leu Asp Asp Val Trp  
1 5 10

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Ser Arg Phe Ile Ile Thr Ser Arg  
1 5

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Ser Arg Ile Ile Ile Thr Thr Arg  
1 5

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid

- 117 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Ser Arg Ile Ile Leu Thr Thr Arg  
1 5

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Thr Thr Arg  
1

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Gly Leu Pro Leu Thr Leu Lys Val  
1 5

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Gly Leu Pro Leu Ala Leu Lys Val  
1 5

- 118 -

## (2) INFORMATION FOR SEQ ID NO:124:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Gly Leu Pro Leu Ser Val Val Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:125:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Gly Leu Pro Leu Ala Leu Ile Thr  
1 5

## (2) INFORMATION FOR SEQ ID NO:126:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Lys Ile Ser Tyr Asp Ala Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:127:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein



- 119 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Lys Ile Ser Tyr Asp Gly Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:128:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Gly Phe Ser Tyr Lys Asn Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:129:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Val Phe Leu Ser Phe Arg Gly  
1 5

## (2) INFORMATION FOR SEQ ID NO:130:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Pro Ile Phe Tyr Met Val Asp Pro  
1 5

## (2) INFORMATION FOR SEQ ID NO:131:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

- 120 -

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Pro Ile Phe Tyr Asp Val Asp Pro  
1 5

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Val Gly Ile Asp Asp His  
1 5

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

Val Gly Ile Asp Thr His  
1 5

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

Phe Leu Asp Ile Ala Cys Phe  
1 5

(2) INFORMATION FOR SEQ ID NO:135:

- 121 -

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Met His Asp Gln Leu Arg Asp Met Gly  
1 5

- (2) INFORMATION FOR SEQ ID NO:136:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Met His Asp Leu Ile Gln Asp Met Gly  
1 5

- (2) INFORMATION FOR SEQ ID NO:137:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Met His Asp Leu Ile Gln Asp Met Gly  
1 5

- (2) INFORMATION FOR SEQ ID NO:138:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 6 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

- 122 -

Ser Lys Leu Glu Ser Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:139:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Gly Leu His Ser Leu Glu Tyr Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8 base pairs  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Gly Leu Arg Ser Leu Glu Ile Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:141:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3432 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

ACAAGTAAAA GAAAGAGCGA GAAATCATCG AAATGGATTT CATCTCATCT CTTATCGTTG	60
GCTGTGCTCA GGTGTTGTGT GAATCTATGA ATATGGCGGA GAGAAGAGGA CATAAGACTG	120
ATCTTAGACA AGCCATCACT GATCTTGAAA CAGCCATCGG TGAAGCTGAG GCCATACGTG	180
ATGACCTGAC TTTACGGATC CAACAAGACG GTCTAGAGGG ACGAAGCTGC TCRAATCGTG	240
CCAGAGAGTG GCTTAGTGCG GTGCAAGTAA CGGAGACTAA AACAGCCCTA CTTTGTAGTGA	300
GGTTTAGGCG TCGGGAACAG AGGACGCGAA TGAGGAGGAG ATACCTCAGT TGTTTCGGTT	360
GTGCCGACTA CAAACTGTGC AAGAAGGTTT CTGCCATATT GAAGAGCATT GGTGAGCTGA	420
GAGAACGCTC TGAAGCTATC AAAACAGATG GCGGGTCAAT TCAAGTAACT TGTAGAGAGA	480
TACCCATCAA GTCCGTTGTC GGAAATACCA CGATGATGGA ACAGGTTTGT GAATTTCTCA	540

- 123 -

GTGAAGAAGA AGAAGAGGA ATCATTGGTG TTTATGGACC TGGTGGGGTT GGGAAGACAA	600
CGTTAATGCA GAGCATTAAC AACGAGCTGA TCACAAAAGG ACATCAGTAT GATGTACTGA	660
TTTGGGTTCA AATGTCCAGA GAATTCGGCG AGTGTACAAT TCAGCAAGCC GTTGGAGCAC	720
GGTTGGGTTT ATCTTGGGAC GAGAAGGAGA CCGGCGAAAA CAGAGCTTTG AAGATATACA	780
GAGCTTTGAG ACAGAAACGT TTCTTGTGTG TGCTAGATGA GTCTGGGAAG AGATAGACTT	840
GGAGAAAACCT GGAGTTCCTC GACCTTGACA GGGAAAACAA ATGCAAGGTG ATGTTACCGA	900
CACGGTCTAT AGCATTATGC AACAAATATG GTGCGGAATA CAAGTTGAGA GTGGAGTTTC	960
TGGAGAAGAA ACACGCGTGG GAGCTGTTCT GTAGTAAGGT ATGGAGAAAA GATCTTTTAG	1020
AGTCATCATC AATTCGCCGG CTCGCGGAGA TTATAGTGAG TAAATGTGGA GGATTGCCAC	1080
TAGCGTTGAT CACTTTAGGA GGAGCCATGG CTCATAGAGA GACAGAAGAA GAGTGGATCC	1140
ATGCTAGTGA AGTTCTGACT AGATTTCAG CAGAGATGAA GGGTATGAAC TATGTATTTG	1200
CCCTTTTGAA ATTCAGCTAC GACAACCTCG AGAGTGATCT GCTTCGGTCT TGTTCCTTGT	1260
ACTGCGCTTT ATTCCAGAA GAACATTGTA TAGAGATCGA GCAGCTTGTT CAGTACTGGG	1320
TCGGCGAAGG GTTCTCACC AGCTCCCATG GCGTTAACAC CATTTACAAG GGATATTTTC	1380
TCATTGGGGA TCTGAAAGCG GCATGTTTGT TGGAAACCGG AGATGAGAAA ACACAGGTGA	1440
AGATGCATAA TGTGGTCAGA AGCTTTGCAT TGTGGATGGC ATCTGAACAG GGGACTTATA	1500
AGGAGCTGAT CCTAGTTGAG CCTAGCATGG GACATACTGA AGCTCCTAAA GCAGAAAACCT	1560
GGCGACAAGC TTGGTGATCT CATTGTTAGA TAACAGAATC CAGACCTTGC CTGAAAAACCT	1620
CATATGCCCG AAACCTGACAA CACTGATGCT CCAACAGAAC AGCTCTTTGA AGAAGATTCC	1680
AACAGGGTTT TTCATGCATA TGCTGTCTCT CAGAGTCTTG GACTTGTCGT TCACAAGTAT	1740
CACTGAGATT CCGTTGTCTA TCAAGTATTT GGTGGAGTTG TATCATCTGT CTATGTCAGG	1800
AACAAAGATA AGTGATTGTC CACAGGAGCT TGGGAATCTT AGAAAACTGA AGCATCTGGA	1860
CCTACAAAGA ACTCAGTTTC TTCAGACGAT CCCACGAGAT GCCATATGTT GGCTGAGCAA	1920
GCTCGAGGTT CTGAACCTGT ACTACAGTTA CGCCGGTTGG GAAGTGCAGA GCTTTGGAGA	1980
AGATGAAGCA GAAGAACTCG GATTGCTGA CTTGGAATAC TTGGAAAACC TAACCACACT	2040
CGGTATCACT GTTCTCTCAT TGGAGACCCT AAAAATCTC TTCGAGTTTC GTGCTTTGCA	2100
TAAACATATA CAGCATCTCC ACGTTGAAGA GTGCAATGAA CTCCTCTACT TCAATCTCCC	2160
ATCACTCACT AACCATGGCA GGAACCTGAG AAGACTTAGC ATTAAGAGTT GCCATGACTT	2220
GGAGTACCTG GTCACACCCG CAGATTTTGA AAATGATTGG CTTCCGAGTC TAGAGGTTCT	2280
GACGTTACAC AGCCTTCACA ACTTAACCAG AGTGTGGGGA AATTCTGTAA GCCAAGATTG	2340
TCTGCGGAAT ATCCGTTGCA TAAACATTTT AACTGCAAC AAGCTGAAGA ATGTCTCATG	2400
GGTTCAGAAA CTCCCAAAGC TAGAGGTGAT TGAAGTTTC GACTGCAGAG AGATAGAGGA	2460

- 124 -

ATTGATAAGC GAACACGAGA GTCCATCCGT CGAAGATCCA ACATTGTTCC CAAGCCTGAA 2520  
 GACCTTGAGA ACTAGGGATC TGCCAGAACT AAACAGCATC CTCCCATCTC GATTTTCATT 2580  
 CCAAAAAGTT GAAACATTAG TCATCACAAA TTGCCCCAGA GTTAAGAAAC TGCCGTTTCA 2640  
 GGAGAGGAGG ACCCAGATGA ACTTGCCAAC AGTTTATTGT GAGGAGAAAT GGTGGAAAGC 2700  
 ACTGGAAAAA GTTGAAACAT TAGTCATCAC AAATTGCCCC AGAGTTAAGA AACTGCCGTT 2760  
 TCAGGAGAGG AGGACCCAGA TGAACCTGCC AACAGTTTAT TGTGAGGAGA AATGGTGGAA 2820  
 AGCACTGGAA AAAGATCAAC CAAACGAAGA GCTTTGTTAT TTACCGCGCT TTGTTCCAAA 2880  
 TTGATATAAG AGCTAAGAGC ACTCTGTACA AATATGTCCA TTCATAAGTA GCAGGAAGCC 2940  
 AGGAAGGTTG TTCCAGTGAA GTCATCAACT TTCCACTAGA CCACAAAACCT AGAGATTATG 3000  
 TAATCATAAA AACCAACTA TCCGCGATCA AATAGATCTC ACGACTATGA GGACGAAGAC 3060  
 TCACCGAGTA TCGTCGATAT AGAAACTCCA AGCTCCAGTT CCGATCAGTG AAGACGAACA 3120  
 AGTTTATCAG ATCTCTGCAA CAATTCTGGG AATCGTCACC TCAGATTAGA CCTCCAGTAA 3180  
 GAAGTGAGAA AGCATGGACG ACGACTGTGA AGAATTGAGC TAATGAGCTG AACCGGATCC 3240  
 GGTGAAATTG CAGAACCGGA TCGGAGAAGA AGAATTTTGC ATTTGTGCAT CTTTATTTTT 3300  
 AATTGTTACG TTTGAGCCCC AATAATCATA GATATTGTAG TGAAGACCAA ATTCATGGT 3360  
 GGATCAATCA AATTGTATTT TCAAATTTTC GTAGTGTAAT AACGGAAAAA GGAATAAAAA 3420  
 GGTCACTGAG TA 3432

## (2) INFORMATION FOR SEQ ID NO:142:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 909 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Met Asp Phe Ile Ser Ser Leu Ile Val Gly Cys Ala Gln Val Leu Cys  
 1 5 10 15  
 Glu Ser Met Asn Met Ala Glu Arg Arg Gly His Lys Thr Asp Leu Arg  
 20 25 30  
 Gln Ala Ile Thr Asp Leu Glu Thr Ala Ile Gly Asp Leu Lys Ala Ile  
 35 40 45  
 Arg Asp Asp Leu Thr Leu Arg Ile Gln Gln Asp Gly Leu Glu Gly Arg  
 50 55 60  
 Ser Cys Ser Asn Arg Ala Arg Glu Trp Leu Ser Ala Val Gln Val Thr  
 65 70 75 80  
 Glu Thr Lys Thr Ala Leu Leu Leu Val Arg Phe Arg Arg Arg Glu Gln

85

90

95

Arg	Thr	Arg	Met	Arg	Arg	Arg	Tyr	Leu	Ser	Cys	Phe	Gly	Cys	Ala	Asp
			100					105						110	
Tyr	Lys	Leu	Cys	Lys	Lys	Val	Ser	Ala	Ile	Leu	Lys	Ser	Ile	Gly	Glu
		115					120					125			
Leu	Arg	Glu	Arg	Ser	Glu	Ala	Ile	Lys	Thr	Asp	Gly	Gly	Ser	Ile	Gln
	130					135					140				
Val	Thr	Cys	Arg	Glu	Ile	Pro	Ile	Lys	Ser	Val	Val	Gly	Asn	Thr	Thr
145					150					155					160
Met	Met	Glu	Gln	Val	Leu	Glu	Phe	Leu	Ser	Glu	Glu	Glu	Glu	Arg	Gly
				165					170					175	
Ile	Ile	Gly	Val	Tyr	Gly	Pro	Gly	Gly	Val	Gly	Lys	Thr	Thr	Leu	Met
			180					185					190		
Gln	Ser	Ile	Asn	Asn	Glu	Leu	Ile	Thr	Lys	Gly	His	Gln	Tyr	Asp	Val
		195					200					205			
Leu	Ile	Trp	Val	Gln	Met	Ser	Arg	Glu	Phe	Gly	Glu	Cys	Thr	Ile	Gln
	210					215					220				
Gln	Ala	Val	Gly	Ala	Arg	Leu	Gly	Leu	Ser	Trp	Asp	Glu	Lys	Glu	Thr
225					230					235					240
Gly	Glu	Asn	Arg	Ala	Leu	Lys	Ile	Tyr	Arg	Ala	Leu	Arg	Gln	Lys	Arg
				245					250					255	
Phe	Leu	Leu	Leu	Leu	Asp	Asp	Val	Trp	Glu	Glu	Ile	Asp	Leu	Glu	Lys
			260					265					270		
Thr	Gly	Val	Pro	Arg	Pro	Asp	Arg	Glu	Asn	Lys	Cys	Lys	Val	Met	Phe
		275					280					285			
Thr	Thr	Arg	Ser	Ile	Ala	Leu	Cys	Asn	Asn	Met	Gly	Ala	Glu	Tyr	Lys
	290					295					300				
Leu	Arg	Val	Glu	Phe	Leu	Glu	Lys	Lys	His	Ala	Trp	Glu	Leu	Phe	Cys
305					310					315					320
Ser	Lys	Val	Trp	Arg	Lys	Asp	Leu	Leu	Glu	Ser	Ser	Ser	Ile	Arg	Arg
				325					330					335	
Leu	Ala	Glu	Ile	Ile	Val	Ser	Lys	Cys	Gly	Gly	Leu	Pro	Leu	Ala	Leu
			340					345					350		
Ile	Thr	Leu	Gly	Gly	Ala	Met	Ala	His	Arg	Glu	Thr	Glu	Glu	Glu	Trp
		355					360					365			
Ile	His	Ala	Ser	Glu	Val	Leu	Thr	Arg	Phe	Pro	Ala	Glu	Met	Lys	Gly
	370					375					380				
Met	Asn	Tyr	Val	Phe	Ala	Leu	Leu	Lys	Phe	Ser	Tyr	Asp	Asn	Leu	Glu
385					390					395					400
Ser	Asp	Leu	Leu	Arg	Ser	Cys	Phe	Leu	Tyr	Cys	Ala	Leu	Phe	Pro	Glu
				405					410					415	
Glu	His	Ser	Ile	Glu	Ile	Glu	Gln	Leu	Val	Glu	Tyr	Trp	Val	Gly	Glu

- 126 -

420							425					430				
Gly	Phe	Leu	Thr	Ser	Ser	His	Gly	Val	Asn	Thr	Ile	Tyr	Lys	Gly	Tyr	
		435					440					445				
Phe	Leu	Ile	Gly	Asp	Leu	Lys	Ala	Ala	Cys	Leu	Leu	Glu	Thr	Gly	Asp	
	450					455					460					
Glu	Lys	Thr	Gln	Val	Lys	Met	His	Asn	Val	Val	Arg	Ser	Phe	Ala	Leu	
	465				470					475					480	
Trp	Met	Ala	Ser	Glu	Gln	Gly	Thr	Tyr	Lys	Glu	Leu	Ile	Leu	Val	Glu	
				485					490					495		
Pro	Ser	Met	Gly	His	Thr	Glu	Ala	Pro	Lys	Ala	Glu	Asn	Trp	Arg	Gln	
			500					505					510			
Ala	Leu	Val	Ile	Ser	Leu	Leu	Asp	Asn	Arg	Ile	Gln	Thr	Leu	Pro	Glu	
		515					520					525				
Lys	Leu	Ile	Cys	Pro	Lys	Leu	Thr	Thr	Leu	Met	Leu	Gln	Gln	Asn	Ser	
	530					535					540					
Ser	Leu	Lys	Lys	Ile	Pro	Thr	Gly	Phe	Phe	Met	His	Met	Pro	Val	Leu	
	545				550					555					560	
Arg	Val	Leu	Asp	Leu	Ser	Phe	Thr	Ser	Ile	Thr	Glu	Ile	Pro	Leu	Ser	
				565					570					575		
Ile	Lys	Tyr	Leu	Val	Glu	Leu	Tyr	His	Leu	Ser	Met	Ser	Gly	Thr	Lys	
			580					585					590			
Ile	Ser	Val	Leu	Pro	Gln	Glu	Leu	Gly	Asn	Leu	Arg	Lys	Leu	Lys	His	
		595					600					605				
Leu	Asp	Leu	Gln	Arg	Thr	Gln	Phe	Leu	Gln	Thr	Ile	Pro	Arg	Asp	Ala	
	610					615					620					
Ile	Cys	Trp	Leu	Ser	Lys	Leu	Glu	Val	Leu	Asn	Leu	Tyr	Tyr	Ser	Tyr	
	625				630					635					640	
Ala	Gly	Trp	Glu	Leu	Gln	Ser	Phe	Gly	Glu	Asp	Glu	Ala	Glu	Glu	Leu	
				645					650					655		
Gly	Phe	Ala	Asp	Leu	Glu	Tyr	Leu	Glu	Asn	Leu	Thr	Thr	Leu	Gly	Ile	
			660					665					670			
Thr	Val	Leu	Ser	Leu	Glu	Thr	Leu	Lys	Thr	Leu	Phe	Glu	Phe	Gly	Ala	
		675					680					685				
Leu	His	Lys	His	Ile	Gln	His	Leu	His	Val	Glu	Glu	Cys	Asn	Glu	Leu	
	690					695					700					
Leu	Tyr	Phe	Asn	Leu	Pro	Ser	Leu	Thr	Asn	His	Gly	Arg	Asn	Leu	Arg	
	705				710					715					720	
Arg	Leu	Ser	Ile	Lys	Ser	Cys	His	Asp	Leu	Glu	Tyr	Leu	Val	Thr	Pro	
				725					730					735		
Ala	Asp	Phe	Glu	Asn	Asp	Trp	Leu	Pro	Ser	Leu	Glu	Val	Leu	Thr	Leu	
			740					745					750			
His	Ser	Leu	His	Asn	Leu	Thr	Arg	Val	Trp	Gly	Asn	Ser	Val	Ser	Gln	



- 127 -

755						760						765					
Asp	Cys	Leu	Arg	Asn	Ile	Arg	Cys	Ile	Asn	Ile	Ser	His	Cys	Asn	Lys		
	770						775					780					
Leu	Lys	Asn	Val	Ser	Trp	Val	Gln	Lys	Leu	Pro	Lys	Leu	Glu	Val	Ile		
	785				790					795					800		
Glu	Leu	Phe	Asp	Cys	Arg	Glu	Ile	Glu	Glu	Leu	Ile	Ser	Glu	His	Glu		
				805					810					815			
Ser	Pro	Ser	Val	Glu	Asp	Pro	Thr	Leu	Phe	Pro	Ser	Leu	Lys	Thr	Leu		
			820					825					830				
Arg	Thr	Arg	Asp	Leu	Pro	Glu	Leu	Asn	Ser	Ile	Leu	Pro	Ser	Arg	Phe		
		835					840					845					
Ser	Phe	Gln	Lys	Val	Glu	Thr	Leu	Val	Ile	Thr	Asn	Cys	Pro	Arg	Val		
	850						855					860					
Lys	Lys	Leu	Pro	Phe	Gln	Glu	Arg	Arg	Thr	Gln	Met	Asn	Leu	Pro	Thr		
	865				870					875					880		
Val	Tyr	Cys	Glu	Glu	Lys	Trp	Trp	Lys	Ala	Leu	Glu	Lys	Asp	Gln	Pro		
				885					890					895			
Asn	Glu	Glu	Leu	Cys	Tyr	Leu	Pro	Arg	Phe	Val	Pro	Asn					
			900					905									

(2) INFORMATION FOR SEQ ID NO:143:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

Pro Lys Ala Glu Asn Trp Arg Gln Ala Leu Val Ile Ser Leu Leu Asp  
1 5 10 15

Asn Arg Ile Gln Thr Leu  
20

(2) INFORMATION FOR SEQ ID NO:144:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

- 128 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Pro Glu Lys Leu Ile Cys Pro Lys Leu Thr Thr Leu Met Leu Gln Gln  
 1 5 10 15  
 Asn Ser Ser Leu Lys Lys Ile  
 20

## (2) INFORMATION FOR SEQ ID NO:145:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 24 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

Pro Thr Gly Phe Phe Met His Met Pro Val Leu Arg Val Leu Asp Leu  
 1 5 10 15  
 Ser Phe Thr Ser Ile Thr Glu Ile  
 20

## (2) INFORMATION FOR SEQ ID NO:146:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 23 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

Pro Leu Ser Ile Lys Tyr Leu Val Glu Leu Tyr His Leu Ser Met Ser  
 1 5 10 15  
 Gly Thr Lys Ile Ser Val Leu  
 20

## (2) INFORMATION FOR SEQ ID NO:147:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 24 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

- 129 -

Pro Gln Glu Leu Gly Asn Leu Arg Lys Leu Lys His Leu Asp Leu Gln  
 1 5 10 15  
 Arg Thr Gln Phe Leu Gln Thr Ile  
 20

## (2) INFORMATION FOR SEQ ID NO:148:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 37 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

Pro Arg Asp Ala Ile Cys Trp Leu Ser Lys Leu Glu Val Leu Asn Leu  
 1 5 10 15  
 Tyr Tyr Ser Tyr Ala Gly Trp Glu Leu Gln Ser Phe Gly Glu Asp Glu  
 20 25 30  
 Ala Glu Glu Leu Gly  
 35

## (2) INFORMATION FOR SEQ ID NO:149:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 25 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

Phe Ala Asp Leu Glu Tyr Leu Glu Asn Leu Thr Thr Leu Gly Ile Thr  
 1 5 10 15  
 Val Leu Ser Leu Glu Thr Leu Lys Thr  
 20 25

## (2) INFORMATION FOR SEQ ID NO:150:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 27 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 130 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

Leu Phe Glu Phe Gly Ala Leu His Lys His Ile Gln His Leu His Val  
1 5 10 15  
Glu Glu Cys Asn Glu Leu Leu Tyr Phe Asn Leu  
20 25

## (2) INFORMATION FOR SEQ ID NO:151:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

Pro Ser Leu Thr Asn His Gly Arg Asn Leu Arg Arg Leu Ser Ile Lys  
1 5 10 15  
Ser Cys His Asp Leu Glu Tyr Leu Val Thr  
20 25

## (2) INFORMATION FOR SEQ ID NO:152:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

Pro Ala Asp Phe Glu Asn Asp Trp Leu Pro Ser Leu Glu Val Leu Thr  
1 5 10 15  
Leu His Ser Leu His Asn Leu Thr Arg Val Trp Gly Asn  
20 25

## (2) INFORMATION FOR SEQ ID NO:153:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 131 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

Ser Val Ser Gln Asp Cys Leu Arg Asn Ile Arg Cys Ile Asn Ile Ser  
 1 5 10 15  
 His Cys Asn Lys Leu Lys Asn Val Ser Trp Val Gln Lys Leu  
 20 25 30

## (2) INFORMATION FOR SEQ ID NO:154:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 28 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

Pro Lys Leu Glu Val Ile Glu Leu Phe Asp Cys Arg Glu Ile Glu Glu  
 1 5 10 15  
 Leu Ile Ser Glu His Glu Ser Pro Ser Val Glu Asp  
 20 25

## (2) INFORMATION FOR SEQ ID NO:155:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 22 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Pro Thr Leu Phe Pro Ser Leu Lys Thr Leu Arg Thr Arg Asp Leu Pro  
 1 5 10 15  
 Glu Leu Asn Ser Ile Leu  
 20

## (2) INFORMATION FOR SEQ ID NO:156:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 23 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

- 132 -

Pro Ser Arg Phe Ser Phe Gln Lys Val Glu Thr Leu Val Ile Thr Asn  
 1 5 10 15

Cys Pro Arg Val Lys Lys Leu  
 20

## (2) INFORMATION FOR SEQ ID NO:157:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5134 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

AAGCTTTACA GATTGGATGA TCTCTTAATG CATGCTGAAG TGA CTGCAAA AAGGTTAGCA	60
ATATTCAGTG GTTCTCGTTA TGAATATTTT ATGAACGGAA GCAGCACTGA GAAAATGAGG	120
CCCTTGTTAT CTGATTTTCT GCAAGAGATT GAGTCTGTCA AGGTAGAGTT CAGAAATGTT	180
TGCTTGCAAG TTCTGGATAT ATCACCTTTT TCCCTGACAG ATGGAGAAGG CTTGTTAAT	240
TTCTTATTAA AAAACCAGGC CAAGGTGCCG AATGATGATG CTGTTTCTTC TGATGGAAGT	300
TTAGAGGATG CAAGCAGCAC TGAGAAAATG GGACTTCCAT CTGATTTTCT CCGAGAGATT	360
GAGTCTGTTG AGATAAAGGA GGCCAGAAAA TTATATGATC AAGTTTGGGA TGCAACACAT	420
TGTGAGACGA GTAAGCACGA TGGAAAAAGC TTTATCAACA TTATGTTAAC CCAACAGGAC	480
AAGGTGCTGG ACTATGATGC TGGTTCAGTG TCTTATCTTC TTAACCAAAT CTCAGTAGTT	540
AAAGACAAAA TATTGCACAT TGGCTCTTTA CTTGTAGATA TTGTACAGTA CCGGAATATG	600
CATATAGAAC TTACAGATCT CGCTGAACGT GTTCAAGATA AAAACTACAT TCGTTTCTTC	660
TCTGTCAAGG GTTATATTCC TGCTTGGTAT TACACACTAT ATCTCTCTGA TGTCAAGCAA	720
TTGCTTAAGT TTGTTGAGGC AGAGGTAAAG ATTATTTGTC TGAAAGTACC AGATTCTTCA	780
AGTTATAGCT TCCCTAAGAC AAATGGATTA GGATATCTCA ATTGCTTTTT AGGCAAATTG	840
GAGGAGCTTT TACGTTCTAA GCTCGATTG ATAATCGACT TAAAACATCA GATTGAATCA	900
GTCAAGGAGG GCTTATTGTG CCTAAGATCA TTCATTGATC ATTTTTCAGA AAGCTATGTT	960
GAGCATGATG AAGCTTGTGG TCTTATAGCA AGAGTTTCTG TAATGGCATA CAAGGCTGAG	1020
TATGTCATTG ACTCATGCTT GGCCTATTCT CATCCACTCT GGTACAAAGT TCTTTGGATT	1080
TCTGAAGTTC TTGAGAATAT TAAGCTTGTA AATAAAGTTG TTGGGGAGAC ATGTGAAAGA	1140
AGGAACACTG AAGTTACTGT GCATGAAGTT GCAAAGACTA CCACTAATGT AGCACCATCT	1200
TTTTCAGCTT ATACTCAAAG AGCAAACGAA GAAATGGAGG GTTTTCAGGA TACAATAGAT	1260

- 133 -

GAATTAAAGG ATAAACTACT TGGAGGATCA CCTGAGCTTG ATGTCATCTC AATCGTTGGC	1320
ATGCCAGGAT TGGGCAAGAC TACACTAGCA AAGAAGATT ACAATGATCC AGAAGTCACC	1380
TCTCGCTTCG ATGTCCATGC TCAATGTGTT GTGACTCAAT TATATTCATG GAGAGAGTTG	1440
TTGCTCACCA TTTTGAATGA TGTGCTTGAG CCTTCTGATC GCAATGAAAA AGAAGATGGA	1500
GAAATAGCTG ATGATCTACG CCGATTTTGT TTGACCAAGA GATTCTTGAT TCTCATTGAT	1560
GATGTGTGGG ACTATAAAGT GTGGGACAAT CTATGTATGT GCTTCAGTGA TGTTTCAAAT	1620
AGGAGTAGAA TTATCCTAAC AACCCGCTTG AATGATGTCG CCGAATATGT CAAATGTGAA	1680
AGTGATCCCC ATCATCTTCG TTTATTCAGA GATGACGAGA GTTGGACATT ATTACAGAAA	1740
GAAGTCTTTC AAGGAGAGAG CTGTCCACCT GAACTTGAAG ATGTGGGATT TGAAATATCA	1800
AAAAGTTGTA GAGGGTTGCC TCTCTCAGTT GTGTTAGTAG CTGGTGTCTT GAAACAGAAA	1860
AAGAAGACAC TAGATTCATG GAAAGTAGTA GAACAAAGTC TAAGTTCCCA GAGGATTGGC	1920
AGCTTGGAAG AGAGCATATC TATAATTGGA TTCAGTTACA AGAATTACC AACTATCTT	1980
AAGCCTTGTT TTCTCTATTT TGGAGGATT TTGCAGGGAA AGGATATTCA TGACTCAAAA	2040
ATGACCAAGT TGTGGGTAGC TGAAGAGTTT GTACAAGCAA ACAACGAAAA AGGACAAGAA	2100
GATACCCGCA CAAGGTTTCT TGGACGATCT TATTGGTAGG AATCTGGTGA TGGCCATGGA	2160
GAAGAGACCT AATGCCAAGG TGA AACGTG CCGCATT CAT GATTGTGTC ATAAATTCTG	2220
CATGGAAAAG GCCAAACAAG AGGATTTCCT TCTCCAGATC AATAGGTAAA AAAA ACTGTA	2280
TTAATTTTAC ATTACAAAAA AAAAGAACTG TATTAATTTT ACTGTATTAT GTTTATGCCA	2340
ACTCTCATTT CCATGTGTTT TCTTTTATTC AATTCAGTGG AGAAGGTGTA TTTCTGAAC	2400
GATTGGAAGA ATACCGATTG TTCGTTTATT CTTACCAAGA TGAAATTGAT CTGTGGCGCC	2460
CATCTCGCTC TAATGTCCGC TCTTTACTAT TCAATGCAAT TGATCCAGAT AACTTGTTAT	2520
GGCCGCGTGA TATCTCCTTC ATTTTGTAGA GCTTCAAGCT TGTAAAGTG TTGGATTGG	2580
AATCATTCAA CATTGGTGGT ACTTTTCCCA TTGAAACACA ATATCTAATT CAGATGAAGT	2640
ACTTTGCGGC CCAAACTGAT GCAAATCAA TTCCTTCATC TATAGCTAAG CTTGAAAATC	2700
TTGAGACTTT TGTGTAAGA GGATTGGGAG GAGAGATGAT ATTACCTTGT TCACTTCTGA	2760
AGATGGTGAA ATTGAGGCAT ATACATGTAA ATGATCGGGT TTCTTTTGGT TTGCGTGAGA	2820
ACATGGATGT TTAACTGGT AACTCACAAT AACCTAATT GGAAACCTTT TCTACTCCGC	2880
GTCTCTTTTA TGGTAAAGAC GCAGAGAAGA TTTTGAGGAA GATGCCAAAA TTGAGAAAAT	2940
TGAGTTGCAT ATTTTCAGGG ACATTGGTT ATTCAAGGAA ATTGAAGGGT AGGTGTGTTT	3000
GTTTTCCAG ATTAGATTTT CTAAGTCACC TTGAGTCCCT CAAGCTGGTT TCGAACAGCT	3060
ATCCAGCCAA ACTTCCTCAC AAGTTCAATT TCCCCTCGCA ACTAAGGGAA CTGACTTTAT	3120
CAAAGTTCCG TCTACCTTGG ACCCAAATTT CGATCATTCG AGAACTGCCC AACTTGGTGA	3180

- 134 -

TTCTTAAGTT ATTGCTCAGA GCCTTTGAAG GGGATCACTG GGAAGTGAAA GATTCAGAGT	3240
TCCTAGAACT CAAATACTTA AAACCTGGACA ACCTCAAAGT TGTACAATGG TCCATCTCTG	3300
ATGATGCTTT TCCTAAGCTT GAACATTTGG TTTTAACGAA ATGTAAGCAT CTTGAGAAAA	3360
TCCCTTCTCG TTTTGAAGAT GCTGTTTGTC TAAATAGAGT TGAGGTGAAC TGGTGCAACT	3420
GGAATGTTGC CAATTCAGCC CAAGATATTC AAACCTATGCA ACATGAAGTT ATAGCAAATG	3480
ATTCATTCAC AGTACTATA CAGCCTCCAG ATTGGTCTAA AGAACAGCCC CTTGACTCTT	3540
AGCAAAGGTT TGTCTTGCT GTGTTTCATCC AAGTGCATTT AACATTTATT CATTTTGTTT	3600
TACACCAGAA CATGTTTATT TTGCTAGTAT TACTTGATAC ATTAAAAGAA ATCGAACTCA	3660
TATTTCTGCT ACAGTCTTAA CTTTCTTGG GCTTACTTGA GGTCTAGATT AGATCAATGG	3720
TTCATGTAAT TTTAATTCA CTGTTTCATT CAACTGTCTT ATGATAGTTG TGAAATGACA	3780
ATATTGTTAT CCCTAGCCAA ATTTATTATG TTCAAATGAA AACTGATGTC ACAACTACTT	3840
TTTTGTGAAA TGTTTTTGAA TTTTTTGCTA TAAAATTGAC GAATTGACAG CTTCTATATT	3900
TGTCAGCTAA ACTCTTTGTC ACCAGAAGTG TATTTAGAAT TACTGTGGTT TTATGAAAGA	3960
GTTCTGTAGA ATTTATGCT TTTGCAGAAT ATAGTTTAAA ACAACAACAC TTCTCTGTTT	4020
CAGAGATAGC AGAAGCTAAA GTTCAAGGCA TTTTGTTTAT TTCTAGAACA AGTGGAGTTC	4080
TTATGTTGAA TTCTTGAAAA GAAGAAGAAT CAGGAGCAGG TAAAGTTATC TCTTTTATG	4140
TTTTTCTTCT TTAGATGTT ATTTCTTCAT CTTGAACGTG AACACCGCTG AAAGCATTTT	4200
AATAAAACCG GAGAGAAAAA TAAGATCTTT TTATATAAAG CATTATCATG TAAATATGCC	4260
TAAATCCATA TGGTACAACT GTTTGACAAA ATGATAGAGA GGGGAGTTT ATAGTATAAG	4320
TAAAACAGGA TTGAGAAAAA AATCCTTGCA CGATTTTCAA TTTCTGGCCA CATCACAATG	4380
TGTGTCAAAG TCCCCCTCTT TAAGTGGAAC AAGCAATCAG AAAAGCTCAT TCTTATCGGT	4440
GACATACCAA TACCAGCTGA CTGTCTCATC TTGGTTAACT TAGCCTTGCT TACTTAGACT	4500
ATTAGATTAG TTAATAATGA ACTGGTAAAT TGGAACCAA TGTAGTTAGC TTGATGAGCT	4560
GGTAGACATG TATATATGAA GATACACGCG TAACTTTAGT CGATGGTTAA TTTTTCATTT	4620
TTGATTTTTT TTCTTCACAG AGTATATATG AACTTGGCCT AAAAGTTTGT CTTCACTAAT	4680
TTAACTATTA CCGTGGATGA AACAAGCATG GCAACATTTT CAACAACAT CACTCAAGCA	4740
ATGTAAAAA TGGAGGTTCT ACGAGCGGTA CATGTAAGAG TTTTGTGCAC ACAAGAGGTT	4800
CTGAGACTTG AACCATCCAT GTCCAAGGCA GTTGAGATGC TAGTAAAGAA AGAAGAAGAT	4860
GAGCCTGCAC TAATTAATCT CCCTGTATGA ATGAGAGAAT GAGAAAAAGA TGGAGCTTCA	4920
TGAACCAAAA GTTACCTTTT TTTTTCTTC TTAATGGCAT TACTTTGAAG CACATGTTTG	4980
TTAGTTGTAA ATTGTAATGG TGAAGTGTTT GTAAATATAG GGAGTGATAT TTGAAAGAAT	5040
GGTTGTGTTA TCTTTACAAA CCGGAATCAT TTCTGTATAA TTTTCTTCTG TAATTTTGG	5100



- 135 -

TTTCGGTTTA TTCATTACTC ATTTTCAGTAA GCTT

5134

## (2) INFORMATION FOR SEQ ID NO:158:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

GGNATGGGNG GNNTNGGNAA RACNAC

26

## (2) INFORMATION FOR SEQ ID NO:159:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

NCGNGWNGTN AKDAWNCGNA

20

## (2) INFORMATION FOR SEQ ID NO:160:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

GGWNTBGGWA ARACHAC

17

## (2) INFORMATION FOR SEQ ID NO:161:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 136 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

NRYNRDNGTN GTYTTNCCNA NNCCNNSNRK NCC

33

(2) INFORMATION FOR SEQ ID NO:162:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

GGNMYNSSNG GNNTNGGNAA RACNAC

26

(2) INFORMATION FOR SEQ ID NO:163:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

TYGAYGAYRT BRA

13

(2) INFORMATION FOR SEQ ID NO:164:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

TYCCAAYRT CRTCA

16

(2) INFORMATION FOR SEQ ID NO:165:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid

- 137 -

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

VYMNAYRTCR TCNADNAVNA NNARNA

26

(2) INFORMATION FOR SEQ ID NO:166:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

WWNMRRDTNY TNNTNBTNHT NNARNA

26

(2) INFORMATION FOR SEQ ID NO:167:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

NCGNGWNGTN AKDWNCGNGA

20

(2) INFORMATION FOR SEQ ID NO:168:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

NCKNSWNGTN ADDATDAATN G

21

- 138 -

## (2) INFORMATION FOR SEQ ID NO:169:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

NARNGGNARN CC

12

## (2) INFORMATION FOR SEQ ID NO:170:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

GGWYTBCCWY TBGCHYT

17

## (2) INFORMATION FOR SEQ ID NO:171:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

ARDGCVARWG GVARNCC

17

## (2) INFORMATION FOR SEQ ID NO:172:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 139 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

NRNNWYNAVN SHNARNGGNA RNCC

24

## (2) INFORMATION FOR SEQ ID NO:173:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

GGNYTNCCNY TNDSNBT

17

## (2) INFORMATION FOR SEQ ID NO:174:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

ARRTTRTCRT ADSWRAWYTT

20

## (2) INFORMATION FOR SEQ ID NO:175:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

ARNYYNTYRT ANSRNANNYY

20

## (2) INFORMATION FOR SEQ ID NO:176:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- 140 -

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

RRNWITHWSNT AYRANRVNY

19

(2) INFORMATION FOR SEQ ID NO:177:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

GTNTTYTNTW SNTTYMGRGG

20

(2) INFORMATION FOR SEQ ID NO:178:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

CCNATHTTYT AYRWBGTINGA YCC

23

(2) INFORMATION FOR SEQ ID NO:179:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

GTNGGNATHG AYRMNCA

17

(2) INFORMATION FOR SEQ ID NO:180:

- (i) SEQUENCE CHARACTERISTICS:

- 141 -

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

RAARCANGCD ATRTCNARRA A

21

(2) INFORMATION FOR SEQ ID NO:181:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

TTYTNGAYA THGNTGYTT

20

(2) INFORMATION FOR SEQ ID NO:182:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

CCCATCYK NADNWRRTCR TGCAT

25

(2) INFORMATION FOR SEQ ID NO:183:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

- 142 -

ATGCAYGAYY WNHTNMRRGA YATGGG

26

## (2) INFORMATION FOR SEQ ID NO:184:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

NARNSWYTYN ARYTT

15

## (2) INFORMATION FOR SEQ ID NO:185:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

WSNAARYTNR ARWSNYT

17

## (2) INFORMATION FOR SEQ ID NO:186:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

DWWYTCNARN SWNYKNARN C

21

## (2) INFORMATION FOR SEQ ID NO:187:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA



- 143 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

GGNYTNMRNW NNYTNGA

17

(2) INFORMATION FOR SEQ ID NO:188

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

Leu Lys Phe Ser Tyr Asp Asn Leu Glu Ser Asp Leu Leu  
1                      5                      10

(2) INFORMATION FOR SEQ ID NO:189:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

Gly Val Tyr Gly Pro Gly Gly Val Gly Lys Thr Thr Leu Met Gln Ser  
1                      5                      10                      15

(2) INFORMATION FOR SEQ ID NO:190:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Gly Gly Leu Pro Leu Ala Leu Ile Thr Leu Gly Gly Ala Met

(2) INFORMATION FOR SEQ ID NO:191:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid

- 144 -

- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /note= "Xaa is Met or Pro"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /note= "Xaa is Gly or Pro"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "Xaa is Ile, Leu or Val"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /note= "Xaa is Ile, Leu or Thr"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /note= "Xaa is Ala or Met"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

Gly Xaa Xaa Gly Xaa Gly Lys Thr Thr Xaa Xaa  
1 5 10

(2) INFORMATION FOR SEQ ID NO:192:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /note= "Xaa is Phe or Lys"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /note= "Xaa is Arg or Lys"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /note= "Xaa is Ile, Val or Phe"

- 145 -

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "Xaa is Ile, Leu or Val"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /note= "Xaa is Ile or Leu"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /note= "Xaa is Ile or Val"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /note= "Xaa is Ile, Leu or Val"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 11
- (D) OTHER INFORMATION: /note= "Xaa is Asp or Trp"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

Xaa Xaa Xaa Leu Xaa Xaa Xaa Asp Asp Xaa Xaa  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:193:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /note= "Xaa is Ser or Cys"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /note= "Xaa is Arg or Lys"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /note= "Xaa is Phe, Ile or Val"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /note= "Xaa is Ile or Met"

## (ix) FEATURE:

- 146 -

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "Xaa is Ile, Leu or Phe"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /note= "Xaa is Ser, Cys or Thr"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

Xaa Xaa Xaa Xaa Xaa Thr Xaa Arg  
1 5

## (2) INFORMATION FOR SEQ ID NO:194:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "Xaa is Thr, Ala or Thr"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /note= "Xaa is Leu or Val"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /note= "Xaa is Ile, Val or Lys"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 8
- (D) OTHER INFORMATION: /note= "Xaa is Val or Thr"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

Gly Leu Pro Leu Xaa Xaa Xaa Xaa  
1 5

## (2) INFORMATION FOR SEQ ID NO:195:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

- 147 -

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /note= "Xaa is Lys or Gly"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /note= "Xaa is Ile or Phe"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "Xaa is Asp or Lys"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /note= "Xaa is Ala, Gly or Asn"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

Xaa Xaa Ser Tyr Xaa Xaa Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO:196:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

Asn Ser His Arg  
1

## (2) INFORMATION FOR SEQ ID NO:197:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

Arg Asp Arg Arg Arg Val Asp Pro Cys  
1 5

## (2) INFORMATION FOR SEQ ID NO:198:

- 148 -

- (1) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 4 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: not relevant  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

Thr Gly Asp Leu  
1

(2) INFORMATION FOR SEQ ID NO:199:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 4 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: not relevant  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

His Gly Thr Tyr  
1

(2) INFORMATION FOR SEQ ID NO:200:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 11 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: not relevant  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

Arg Met Ser His Gly Phe Arg Asn Ser Gln Ser  
1                    5                    10

(2) INFORMATION FOR SEQ ID NO:201:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 27 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: not relevant  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 149 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

Gly	Glu	Met	Val	Glu	Ser	Thr	Gly	Lys	Arg	Ser	Thr	Lys	Arg	Arg	Ala
1				5				10					15		
Leu	Leu	Phe	Thr	Ala	Leu	Cys	Ser	Lys	Leu	Ile					
			20					25							

- 150 -

Claims

1. A substantially pure oligonucleotide comprising the sequence:  
5' GGNATGGGNGGNNTNGGNAARACNAC 3', [SEQ. ID NO:158]  
5 wherein N is A, T, G, or C; and R is A or G.
2. A substantially pure oligonucleotide comprising the sequence:  
5' NARNGGNARNCC 3', [SEQ. ID NO: 169] wherein N is A, T, G or C; and R is A or G.
- 10 3. A substantially pure oligonucleotide comprising the sequence:  
5'NCGNGWNGTNAKDAWNCNGNA 3', [SEQ. ID NO: 159]  
wherein N is A, T, G or C; W is A or T; D is A, G, or T; and K is G or T.
- 15 4. A substantially pure oligonucleotide comprising the sequence:  
5' GGWNTBGGWAARACHAC 3', [SEQ ID NO: 160] wherein N is A, T, G or C; R is G or A; B is C, G, or T; H is A, C, or T; and W is A or T.
- 20 5. A substantially pure oligonucleotide comprising the sequence:  
5' TYGAYGAYRTBKRBRA 3', [SEQ. ID NO: 163] wherein R is G or A; B is C, G, or T; D is A, G, or T; Y is T or C; and K is G or T.
- 25 6. A substantially pure oligonucleotide comprising the sequence:  
5' TYCCA VAYRTCRTCNA 3', [SEQ ID NO: 164] wherein N is A, T, G or C; R is G or A; V is G or C or A; and Y is T or C.



- 151 -

7. A substantially pure oligonucleotide comprising the sequence:

5' GGWYTBCCWYTBGCHYT 3', [SEQ ID NO.: 170] wherein B is C, G, or T; H is A, C, or T; W is A or T; and Y is T or C.

8. A substantially pure oligonucleotide comprising the sequence:

5' ARDGCVARWGGVARNCC 3', [SEQ ID NO: 171] wherein N is A, T, G or C; R is G or A; W is A or T; D is A, G, or T; and V is G, C, or A.

9. A substantially pure oligonucleotide comprising the sequence:

5' ARRTTRTCRTADSWRAWYTT 3', [SEQ ID NO: 174] wherein R is G or A; W is A or T; D is A, G, or T; S is G or C; and Y is C or T.

10. A recombinant plant gene comprising the DNA sequence:

5' GGNATGGGNGGNNTNGGNAARACNAC 3', [SEQ ID NO: 158] wherein N is A, T, G or C; and R is A or G.

20 11. The gene of claim 10, further comprising the sequence:

5' NARNGGNARNCC 3', [SEQ ID NO: 169] wherein N is A, T, G or C; and R is A or G.

25 12. The gene of claim 11, further comprising the sequence:

5' NCGNGWNGTNAKDAWNCGNGA 3', [SEQ ID NO: 167] wherein N is A, T, G or C; W is A or T; D is A, G or T; and K is G or T.

SUBSTITUTE SHEET (RULE 26)

- 152 -

13. A recombinant plant gene comprising a combination of any two or more sequences of claims 10, 11 and 12.

14. A substantially pure plant polypeptide  
5 comprising the amino acid sequence:

Gly Xaa<sub>1</sub> Xaa<sub>2</sub> Gly Xaa<sub>3</sub> Gly Lys Thr Thr Xaa<sub>4</sub> Xaa<sub>5</sub>,  
[SEQ ID NO: 191], wherein Xaa<sub>1</sub> is Met or Pro; Xaa<sub>2</sub> is Gly  
or Pro; Xaa<sub>3</sub> is Ile, Leu, or Val; Xaa<sub>4</sub> is Ile, Leu, or  
Thr; and Xaa<sub>5</sub> is Ala or Met.

10 15. A substantially pure plant polypeptide  
comprising the amino acid sequence:

Xaa<sub>1</sub> Xaa<sub>2</sub> Xaa<sub>3</sub> Leu Xaa<sub>4</sub> Xaa<sub>5</sub> Xaa<sub>6</sub> Asp Asp Xaa<sub>7</sub>  
Xaa<sub>8</sub>, [SEQ ID NO: 192],  
wherein Xaa<sub>1</sub> is Phe or Lys; Xaa<sub>2</sub> is Arg or Lys; Xaa<sub>3</sub> is  
15 Ile, Val, or Phe; Xaa<sub>4</sub> is Ile, Leu, or Val; Xaa<sub>5</sub> is Ile or  
Leu; Xaa<sub>6</sub> is Ile or Val; Xaa<sub>7</sub> is Ile, Leu, or Val; and  
Xaa<sub>8</sub> is Asp or Trp.

16. A substantially pure plant polypeptide  
comprising the amino acid sequence:

20 Xaa<sub>1</sub> Xaa<sub>2</sub> Xaa<sub>3</sub> Xaa<sub>4</sub> Xaa<sub>5</sub> Thr Xaa<sub>6</sub> Arg, [SEQ ID NO:  
193]  
wherein Xaa<sub>1</sub> is Ser or Cys; Xaa<sub>2</sub> is Arg or Lys; Xaa<sub>3</sub> is  
Phe, Ile, or Val; Xaa<sub>4</sub> is Ile, or Met; Xaa<sub>5</sub> is Ile, Leu,  
or Phe; Xaa<sub>6</sub> is Ser, Cys, or Thr.

25 17. A substantially pure plant polypeptide  
comprising the amino acid sequence:

Gly Leu Pro Leu Xaa<sub>1</sub> Xaa<sub>2</sub> Xaa<sub>3</sub> Xaa<sub>4</sub>, [SEQ ID NO:  
194],  
wherein Xaa<sub>1</sub> is Thr, Ala, or Ser; Xaa<sub>2</sub> is Leu or Val; Xaa<sub>3</sub>  
30 is Ile, Val, or Lys; and Xaa<sub>4</sub> is Val or Thr.

- 153 -

18. A substantially pure plant polypeptide comprising the amino acid sequence:

Xaa<sub>1</sub> Xaa<sub>2</sub> Ser Tyr Xaa<sub>3</sub> Xaa<sub>4</sub> Leu, {SEQ ID NO: 195},  
wherein Xaa<sub>1</sub> is Lys or Gly; Xaa<sub>2</sub> is Ile or Phe; Xaa<sub>3</sub> is  
5 Asp or Lys; and Xaa<sub>4</sub> is Ala, Gly, or Asn.

19. A method of isolating a disease-resistance gene or fragment thereof from a plant cell, comprising:

(a) providing a sample of plant cell DNA;

(b) providing a pair of oligonucleotides  
10 having sequence homology to a conserved region of an RPS  
disease-resistance gene;

(c) combining said pair of oligonucleotides  
with said plant cell DNA sample under conditions suitable  
for polymerase chain reaction-mediated DNA amplification;  
15 and

(d) isolating said amplified disease-  
resistance gene or fragment thereof.

20. The method of claim 19, wherein said  
amplification is carried out using a reverse-  
20 transcription polymerase chain reaction.

21. The method of claim 19, wherein said reverse-  
transcription polymerase chain reaction is RACE.

22. A method of identifying a plant disease-  
resistance gene in a plant cell, comprising:

25 (a) providing a preparation of plant cell DNA;

(b) providing a detectably-labelled DNA sequence  
having homology to a conserved region of an RPS gene;

(c) contacting said preparation of plant cell DNA  
with said detectably-labelled DNA sequence under  
30 hybridization conditions providing detection of genes  
having 50% or greater sequence identity; and

- 154 -

(d) identifying a disease-resistance gene by its association with said detectable label.

23. The method of claim 22, wherein said DNA sequence is produced according to the method of claim 19.

5        24. The method of claim 22, wherein said preparation of plant cell DNA is isolated from a plant genome.

25. A method of isolating a disease-resistance gene  
10 from a recombinant plant cell library, comprising:  
      (a) providing a recombinant plant cell library;  
      (b) contacting said recombinant plant cell library with a detectably-labelled gene fragment produced according to the method of claim 19 under hybridization  
15 conditions providing detection of genes having 50% or greater sequence identity; and  
      (c) isolating a member of a disease-resistance gene by its association with said detectable label.

26. A method of isolating a disease-resistance  
20 gene from a recombinant plant cell library, comprising:  
      (a) providing a recombinant plant cell library;  
      (b) contacting said recombinant plant cell library with a detectably-labelled oligonucleotide of any of  
claims 1-9 under hybridization conditions providing  
25 detection of genes having 50% or greater sequence identity; and  
      (c) isolating a disease-resistance gene by its association with said detectable label.

- 155 -

27. A recombinant plant polypeptide capable of conferring disease-resistance wherein said plant polypeptide comprises a P-loop domain or nucleotide binding site domain.

5           28. The recombinant plant polypeptide of claim 27, wherein said polypeptide further comprises a leucine-rich repeating domain.

29. A recombinant plant polypeptide capable of conferring disease-resistance wherein said plant  
10 polypeptide contains a leucine-rich repeating domain.

30. A plant disease-resistance gene isolated according to the method comprising:

- (a) providing a sample of plant cell DNA;
- (b) providing a pair of oligonucleotides having  
15 sequence homology to a conserved region of an RPS disease-resistance gene;
- (c) combining said pair of oligonucleotides with said plant cell DNA sample under conditions suitable for polymerase chain reaction-mediated DNA amplification; and  
20 (d) isolating said amplified disease-resistance gene or fragment thereof.

31. A plant disease-resistance gene isolated according to the method comprising:

- (a) providing a preparation of plant cell DNA;
- 25 (b) providing a detectably-labelled DNA sequence having homology to a conserved region of an RPS gene;
- (c) contacting said preparation of plant cell DNA with said detectably-labelled DNA sequence under hybridization conditions providing detection of genes  
30 having 50% or greater sequence identity; and

- 156 -

(d) identifying a disease-resistance gene by its association with said detectable label.

32. A plant disease-resistance gene according to the method comprising:

- 5 (a) providing a recombinant plant cell library;  
(b) contacting said recombinant plant cell library with a detectably-labelled gene fragment produced according to the method of claims 1-4 under hybridization conditions providing detection of genes having 50% or  
10 greater sequence identity; and  
(c) isolating a disease-resistance gene by its association with said detectable label.

33. A method of identifying a plant disease-resistance gene comprising:

- 15 (a) providing a plant tissue sample;  
(b) introducing by biolistic transformation into said plant tissue sample a candidate plant disease-resistance gene;  
(c) expressing said candidate plant disease-  
20 resistance gene within said plant tissue sample; and  
(d) determining whether said plant tissue sample exhibits a disease-resistance response, whereby a response identifies a plant disease-resistance gene.

34. The method of claim 33, wherein said plant  
25 tissue sample comprises leaf, root, flower, fruit, or stem tissue.

35. The method of claim 33, wherein said candidate plant disease-resistance gene is obtained from a cDNA expression library.

- 157 -

36. The method of claim 33, wherein said disease-resistance response is the hypersensitive response.

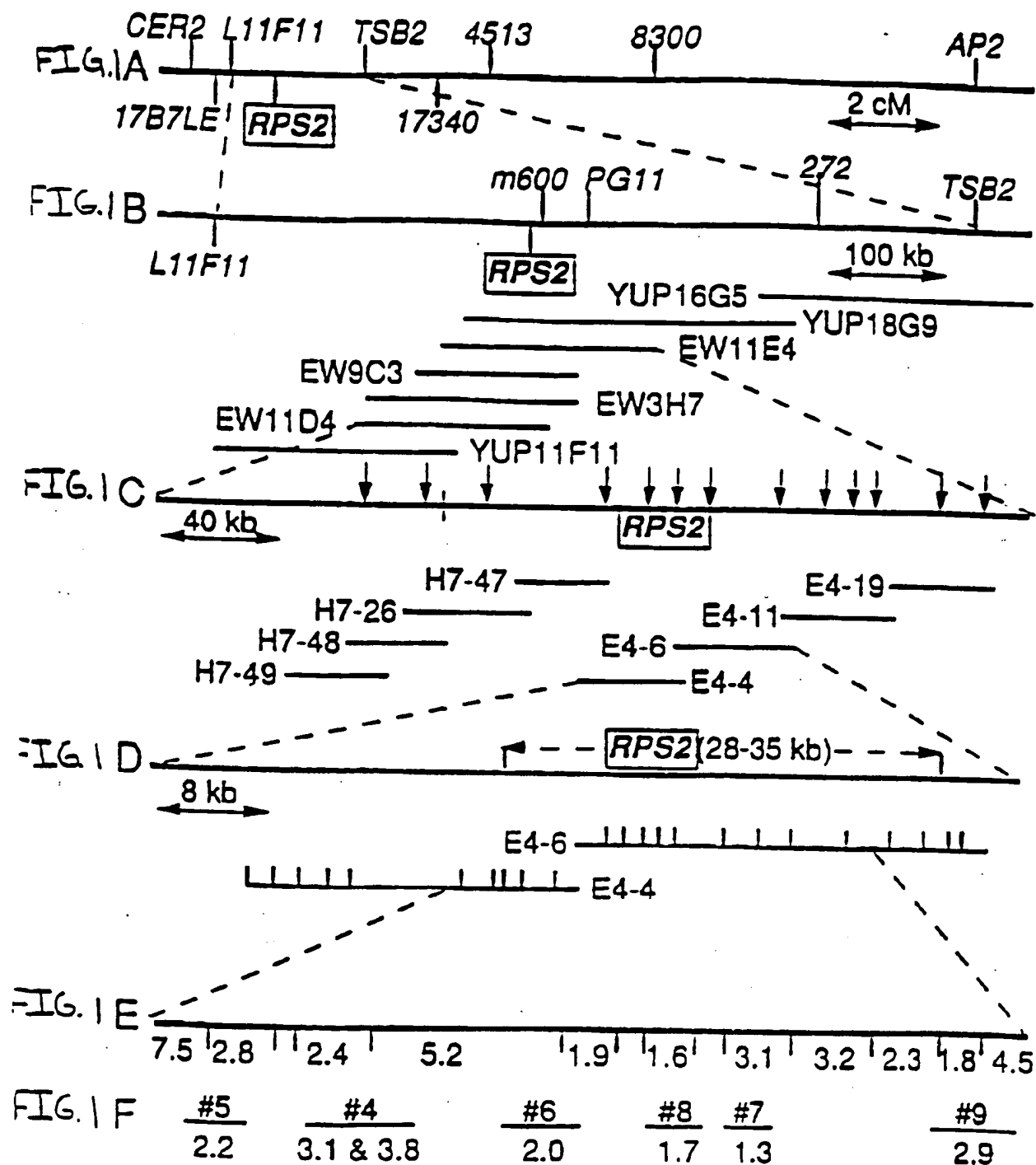
37. A plant disease-resistance gene isolated according to the method comprising:

- 5 (a) providing a plant tissue sample;
- (b) introducing by biolistic transformation into said plant tissue sample a candidate plant disease-resistance gene;
- (c) expressing said candidate plant disease-
- 10 resistance gene within said plant tissue sample; and
- (d) determining whether said plant tissue sample exhibits a disease-resistance response, whereby a response identifies a plant disease-resistance gene.

38. A purified antibody which binds specifically  
15 to an rps family protein.

39. A DNA sequence substantially identical to the DNA sequence shown in Figure 12.

40. A substantially pure polypeptide having a sequence substantially identical to a Prf amino acid  
20 sequence shown in Figure 5 (A or B).





AAGTAAAGAAGGACCGAAGGAATCATTCGAAATGGCAATTCTCATTCTCTTTATCCGTTCGC  
-----+-----  
1 TTEATTTTTCTTTCTCCCTCTTTTAGTAGCTTTACTTAAGTAGAGTAGACAATAGCAAACC 60

a K \* K K E R E I I E M D F I S S L I V C -  
b S K R X S E K S S K W I S S H L L S L A -  
c V K E R A R N H R N G P H L I S Y R W L -

TGTGCTCAGGTGTTTGTCGAATCTATGAATATGCCGGAGACAAGAGGACATAAGACTGAT  
-----+-----  
61 ACACCGAGTCCACAACACACTTAGATACTTATACCCGCTCTCTTCTCTGTATTCTCACTA 120

a C A Q V L C E S M N M A E R R R G B K T D -  
b V L R C C V N L \* I W R R E E D I R L I -  
c C S G V V \* I Y E Y G G E K R T \* D \* S -

CTTAGACAAGCCATCAGTGTATCTTCAAACAGCCATCCGTTGACTTGAAGGCCATACGTTGAT  
-----+-----  
121 GAATCTGTTTCGGTAGTGACTAGAACTTTGTCCGTAGCCACTCAAGTTCCCGGTATCCAATA 180

a L R O A I T D L E T A I G D L X A I R D -  
b L D K P S L I L K Q P S V T \* R P Y V M -  
c \* T S H H \* S \* N S H R \* L E C H T \* . -

CACCTGACTTTACCGATCCAACAGACGGTCTAGACGGAGCAAGCTGCTCAAAATCCCTCCC  
-----+-----  
181 CTGGACTGAAAATCCCTAGGTTGTTCTGCCAGATCTCCCTGCTTCGACCACTTTAGCACCC 240

a D L T L R I Q Q D G L E G R S C S N R A -  
b T \* L Y G S N K T V \* R D E A A Q I V P -  
c P D F T D P T R R S R G T K L L K S C Q -

AGAGACTGGCTTACTGCGGTGCAAGTAACGGAGACTAAAAACAGCCCCACTTTTATGTAGGC  
-----+-----  
241 TCTCTCAGCGAATCAGGCCACGTTTCAATTCCTCTCATTTTGTCCGGATGAAAATCACTCC 300

a R E W L S A V Q V T E T K T A L L L V R -  
b E S G L V R C K \* R R L K Q P Y F \* \* C -  
c R V A \* C G A S N G D \* N S P T F S E V -

TTTAGCGCTCGGCAACAGAGGACCGCAATGACGAGGATACCTCAGTTGTTTCCGTTGT  
-----+-----  
301 AAATCCCGAGCCCTTGTCTCTCTGCCGTTACTCTCTCTATGGAGTCACAAAGCCAAACA 360

a F R R R E Q R T R M R R R Y L S C F G C -  
b L G V G N R G R E \* C G D T S V V S V V -  
c \* A S G T E D A N E E E I P Q \* F R L C -

GCCGACTACAACTGTGCAAGAAGGTTTCTGCCATATTCGAAGAGCATTCGTCAGCTGAG

FIGURE 2

```

361 ----- 420
CGGCTGATGTTTGAACAGCTTCTTCCAAAGACGGTATAAAGTCTCTGTAACCACTGCACTCT
a   A D Y K L C K K V S A I L K S I G E L R -
b   P T T N C A R R F L P Y * R A L V S * E -
c   R L Q T V Q E G F C H I E E H W * A E R -

GAACGCTCTCAACCTATCAAAACAGATGGCCGCTCAATTCAGTAAGTCTCTAGAGAGATA
421 ----- 480
CTTGGGAGACTTCGATAGTTTTGCTACCGCCCACTTAAAGTTCATTGAACATCTCTCTAT
a   E R S E A I K T D G G S I Q V T * R E I -
b   N A L K L S K Q M A G Q F K * L V E R Y -
c   T L * S Y Q N R W R V N S S N L * R D T -

CCCATCAAGTCCGTTCTCGGAAATACCAAGATGATGGAACAGCTTTTGGAAATTTCTAGT
481 ----- 540
CGGTAGTTCAAGCAACAGCCCTTTATGGTGCTACTACCTTGTCCAAAAGCTTAAAGAGTCA
a   P I K S V V C N T T M H E Q V L E P L S -
b   P S S P L S E I P R * W N R F W N P S V -
c   H Q V R C R K Y H D D G T G F G * S Q * -

GAAGAAGAAGAGAGGAATCATTTGTTTATGGACCTGCTGGGCTTGGCAAGACAACG
541 ----- 600
CTTCTTCTTCTTTCTCTTCTAGTAACCAAAATACCTGGACCAACCCCAACCCCTTCTCTTGC
a   E E E E R G I I G V Y G P G G V G K T T -
b   K K K K E E S L V F M D L V G L G R Q R -
c   R R R R K R N H W C L W T W W G W E D N V -

TTAATGCAGAGCATTAACAACGACCTGATCAGAAAAGGACATCAATATGATGTACTGATT
601 ----- 660
AATTACGCTCTGTAATTGTTGCTCGACTAGTGTTTTCTGTAGTCATACTACATGACTAA
a   L M Q S I N N E L I T K G H Q Y D V L I -
b   * C R A L T T S * S Q R D I S H M Y * F -
c   N A E H * Q R A D H K R T S V * C T D L -

TGGGTTCAAAATCTCCAGAGAATTCCGCGAGTGTAATTCAGCAAGCCGTTGGAGCAACG
661 ----- 720
ACCCAAAGTTTACAGGTCTCTTAAGCCGCTTCACATGTTAAGTCTCTCCCAACCTCTGCCC
a   W V Q M S R E F G E C T I Q Q A V G A R -
b   G F K C P E N S A S V Q F S K P L E R G -
c   G S N V Q R I R R V Y N S A S R W S T V -

TTGGGTTTATCTTGGACCGAGAGGACACCGCGAAAACAGAGCTTTTAAAGATATACAGA
721 ----- 780
AACCCAAATAGAACCCTGCTCTTCTCTGCGCCGCTTTTCTCTCGAAATTTCTATATGTCT
a   L G L S W D E K E T G E N R A L K I Y R -
b   W V Y L G T R R R P A K T E L * R Y T E -
c   G F I L G R E G D R R K Q S P E D I Q S -

GCTTTGAGACAGAAACGTTTCTTGTGTTGCTAGATCATGCTCTGGGAAGAGATAGACTTC
781 ----- 840
CGAAACTCTCTCTTTCCAAAGAACAACAACGATCTACTACAGACCTTCTCTATCTGAAC
a   A L R Q X R F L L L L D D V W E E I D L -
b   L * D R N V S C C C * M H S G K R * T W -
c   F E T E T F L V V A R * C L G R D R L C -

```

FIG. 2 CONTINUED

GAGAAAAGTGGAGTTCCTCCACCTGACAGGCGAAAACAAATGCAAGGTGATGTTTCAGGACA  
 841 ----- 900  
 CTCCTTTTGACCTCAAGGAGCTGCACTGTCCCTTTTGTTTACGTTCCACTACAAGTCTGT  
 a E K T G V P R P D R E N K C K V H P T T -  
 b R K L E P L D L T G K T N A R \* C S R H -  
 c E N W S S S T \* Q G K Q M Q G D V H D T -  
 CGGTCTATAGCATTATGCAACAATATGCGGTCCGGAATACAAGTTGAGACTGAGTTTCTC  
 901 ----- 960  
 GCCAGATATCGTAATACGTTGTATATACCCACGCTTATGTTCAACTCTACCTCAAAGAC  
 a R S I A L C N N H G A E Y K L R V E P L -  
 b G L \* H Y A T I W V R N T S \* E W S P W -  
 c V Y S I H Q Q Y G C G I Q V E S G V S G -  
 GAGAAGAAACACGCGTGGGAGCTGTTCTGTAGTAAGGTATGAGAAAAGATCTTTTACAG  
 961 ----- 1020  
 CTCCTCTTTCTGCGCACCCCTCGACAAGACATCATTCATACCTCTTTTGTAGAAAATCTC  
 a E K K H A W E L F C S K V W R K D L L E -  
 b R R N T R G S C S V V R Y G E K I F \* S -  
 c E E T R V G A V L \* \* C M E R R S F R V -  
 TCATCATCAATTCCCGGCTCCGCGAGATTATAGTCAGTAAATGTCGAGCATTGCCACTA  
 1021 ----- 1080  
 AGTAGTAGTTAAGCGGCGGAGCGGCTCTTAATATCACTCATTTACACCTCCTAACCGTAT  
 a S S S I R R L A E I I V S K C G G L P L -  
 b H H Q P A C S R R L \* \* V N V E D C H \* -  
 c I I N S P A R G D Y S E \* M W R I A T S -  
 CGGTGATCACTTTAGGAGCAGCCATCGCTCATAGACACACAGAAGAAGAGTGGATCCAT  
 1081 ----- 1140  
 CGCACTAGTGAATCTCTCTCGGTACCGAGTATCTCTCTGCTTTCTCTACCTAGGTA  
 a A L I T L G G A M A H R E T E E E W I R -  
 b R \* S L \* E E P W L I E R Q R K S G S M -  
 c V D H F R R S H G S \* R D R R R V D P C -  
 CCTAGTGAAGTTCTGACTAGATTTCCAGCAGAGATGAAGGGTATGAACATATGATTTGCC  
 1141 ----- 1200  
 CGATCACTTCAAGACTGATCTAAAGCTGCTCTCTACTTCCCATACTTCATACATAAAGCG  
 a A S E V L T R F P A E H K G M N Y V F A -  
 b L V K P \* L D P Q Q R \* R V \* T M Y L P -  
 c \* \* S S D \* I S S R D E G Y E L C I C P -  
 CTTTGAATTCAGCTACGACAACCTCGAGAGTGATCTGCTTCGGTCTGTTTCTGTAC  
 1201 ----- 1260  
 GAAAACCTTAAAGTCGATGCTGTTGAGCTCTCACTAGACGAAGCCAGAACAAAGACATG  
 a L L K F S Y D N L E S D L L R S C P L Y -  
 b F \* N S A T T T S R V I C F G L V S C T -  
 c F E I Q L R Q P R E \* S A S V L F L V L -  
 TCGGCTTTATCCAGAGAAGACATTCTATAGAGATCGAGCAGCTTGTGAGTACTGGGT  
 1261 ----- 1320  
 ACGCGAAATAGCGGTCTTTCTGTAACATATCTTAGCTCTCGAACAACTCATGACCCAG  
 a C A L F P E E H S I E I E Q L V E Y W V -

FIG. 2 CONTINUED

b A L Y S Q K N I L \* R S S S L L S T G S -  
 c R P I P R R T F Y R D R A A C \* V L G R -  
 -----  
 1321 OCGGAAGGCTTTCTCACCAGCTCCCATGGCGTTAACACCATTTACAAAGCATATTTTCTC 1380  
 CCGCTTCCCAAGAGTGGTTCGAGCGTACCGCAATTGTGTAAATGTTCCCTATAAAAGAG  
 -----  
 a G E G F L T S S H G V N T I Y K G Y F L -  
 b A K G F S P A P M A L T P F T R D I F S -  
 c R R V S H Q L P W R \* H H L Q G I F S H -  
 -----  
 1381 ATTGGGGATCTGAAAGCGGCACTTTGTTGGAAACCGGAGATGAGAAAACACAGCTGAAG 1440  
 TAACCCCTAGACTTTTCGCGGTACAAACAACCTTTGGCGCTCTACTCTTTTGATGTCACCTTC  
 -----  
 a I G D L K A A C L L E T G D E K T Q V K -  
 b L G I \* K R H V C W K P E M R K H R \* R -  
 c W G S E S G M F V G N R R \* E N T G E D -  
 -----  
 1441 ATGCATAATCTGGTCAGAAGCTTTGCCATTGTGGATGGCATCTGAACAGGGGACTTATAAG 1500  
 TAGCTATTACACCACTCTTCCAAACGTAACACCTACCTAGACTTGTCCCTGAAATATTC  
 -----  
 a M H N V V R S F A L W M A S E Q C T Y K -  
 b C I M W S E A L H C G W H L N R G L : R -  
 c A \* C G Q K L C I V D G I \* T G D L \* C -  
 -----  
 1501 GAGCTGATCCTAGTTGAGCCTAGCATGGACATACTGAAGCTCTTAAAGCAGAAAACCTGG 1560  
 CTCGACTAGGATCAACTCCCATCTGACCTGTATGACTTCCAGGATTTCCTCTTTTGACC  
 -----  
 a E L I L V E P S M G H T E A P K A E N W -  
 b S \* S \* L S L A W D I L K L L K Q K T C -  
 c A D P S \* A \* H G T Y \* S S \* S R K L A -  
 -----  
 1561 CGACAAGCGTTGGTGATCTCATTTGTAGATAACAGAATCCAGACCTTGCTGAAAAACTC 1620  
 GCTGTTCCGAACCACTACAGTAACAATCTATTGTCTTAGGTCTGGAACGGACTTTTGTAG  
 -----  
 a R Q A L V I S L L D N R I Q T L P E K L -  
 b D K R W \* S H C \* I T E S R P C L K N S -  
 c T S V G D L I V R \* Q N P D L A \* K T H -  
 -----  
 1621 ATATGCCCGAAAACAGCAACACTGATGCTCCAAACAGAACAGCTCTTTGAAGAACATTCCA 1680  
 TATACCGCGTTTCACTGTTGTGCACTACGAGTTGTCTTGTGAGAAAACCTTTCTTAAGCT  
 -----  
 a I C P K L T T L M L Q Q N S S L K X I P -  
 b Y A R N \* Q H \* C S N R T A L \* R R F Q -  
 c M P E T D N T D A P T E Q L F E E D S N -  
 -----  
 1681 ACAGCGTTTTTCATGCATATGCCCTGTTCTCAGAGTCTTGGACTTGTGTTTCAAGTATC 1740  
 TGTCCCAAAAAGTACGTATACGGACAAGAGTCTCAGAACCTGAACAGCAAGTGTTCATAG  
 -----  
 a T G F F M H M P V L R V L D L S F T S I -  
 b Q G F S C I C L F S E S W T C R S Q V S -  
 c R V F H A Y A C S Q S L G L V V H R Y R -  
 -----  
 1741 ACTGAGATTCCGCTTGTCTATCAAGTATTTGGTGGAGCTGTATCATCTGCTATCTCAGGA 1800  
 TCACTCTAAGGCAACAGATAGTTCTATAAACCACCTCAACATAGTAGACAGATACAGTCTT

FIG. 2 CONTINUED

```

a   T E I P L S I K Y L V E L Y R L S M S G -
b   L R P R C L S S I N W S C I I C L C Q E -
c   * D S V V Y Q V P G G V V S S V Y V R N -

ACAAAGATAAGTGTATTGCCACAGGAGCTTCGGAACTCTTAGAAAACTGAAGCATCTCGAC
1801 ----- 1860
TATTTCTATTACATAACGGGTGTCTCGAACCCTTAGAATCTTTTGACTTCGTAGACCTC

a   T K I S V L P Q E L G N L R K L K H L D -
b   Q R * V Y C H R S L G I L E N * S I W T -
c   K D R C I A T G A W E S * K T E A S G P -

CTACAAAGAACTCAGTTTCTTCAGAGGATCCACAGAGATGCCATATGTTGGCTGAGCAAG
1861 ----- 1920
GATGTTTCTTGAAGTCAAAGAAGTCTGCTAGGGTGCTCTACGGTATACAACCGACTCTTC

a   L Q R T Q F L Q T I P R D A I C W L S K -
b   Y X E L S F F R R S H E M P Y V G * A S -
c   T K N S V S S D D P T R C H M L A E Q A -

CTCGAGGTTCGAAGTGTACTACAGTTACGGCGGTTGSSAACTGCAGAGCTTTGGAGAA
1921 ----- 1980
GAGCTTCAAGACTTGAACATGATGTCAATGCGGCAACCTTTGAGCTCTCGAAACCTCTT

a   L E V L N L Y Y S Y A G W E L Q S F G E -
b   S R P * T C T T V T P V G N C R A L E K -
c   R G S E L V L Q L R R L G T A E L W R R -

GATEAAGCAGAAGAACTCCGATTGCTGACTTGGAACTCTCGAAAAGCTAACCACTC
1981 ----- 2040
CTACTTCTCTTCTTTCAGCCTAAGCGACTGAACCTTATGAACCTTTTGGATTGGTCTGAG

a   D E A E E L G F A D L E Y L E N L T T L -
b   M K Q K N S D S L T W N T W K T * P H S -
c   * S R R T R I R * L G I L G R F N H T R -

GGTATCACTGTTCTCTCATTGAGACCCCTAAAACTCTCTTCGAGTTGGGTGCTTTCCAT
2041 ----- 2100
CCATAGTCACAAGAGAGTAACCTCTCGCATTTTTCAGAGAAGCTCAAGCCACGAAACGTA

a   G I T V L S L E T L R T L F E F G A L H -
b   V S L F S H W R P * K L S S S S V L C I -
c   Y H C S L I G D P K N S L R V R C F A * -

AAACATATACAGCATCTCCAGCTGAAGAGTCCAATGAAGTCTCTTACTTCAATCTCCCA
2101 ----- 2160
TTTGATATATGCTGAGAGCTGCAACTTCTCACGTTACTTGAGGAGATGAAGTTAGACGGT

a   K H I Q H L H V E E C N E L L Y F N L P -
b   N I Y S I S T L K S A M N S S T S I S H -
c   T Y T A S P R * R V Q * T P L L Q S P I -

TCACTCACTAACCATGCCAGCAACCTGAGAAGACTTAGCATTAAGATTGCCATGACTTC
2161 ----- 2220
AGTGAGTGATTCGTACCGTCTCTGAGCTCTTCTGAAATCCTAATTTTCAACGGTACTGAA

a   S L T N H G R N L R R L S I K S C H D L -
b   H S L T M A G T * E D L A L K V A M T W -
c   T H * P W Q E P E R T * H * K L P * L C -

GAGTACCTGCTACACCCGAGATTTTCAAAATGATTCGCTTCGGAGTCTAGAGCTCTC

```

FIG. 2 CONTINUED

2221 ----- 2280  
 CTCATGACCGAGTGTGGCGCTTAAAACCTTTTACTAACCGAAGCCTCAGATCTCCAAAGAC  
 a E Y L V T P A D F E N D W L P S L E V L -  
 b S T W S R P Q I L K H I G F R V R P -  
 c V P G H T R R F K L A S E S R G S D -

2281 ----- 2340  
 ACCTTACACAGCCTTCACAACTTAACCAGAGTGTGGCGAAATTCTGTAAAGCCAAGATTCT  
 TCGAATCTCTCGGAAGCTTTGAATTGGTCTCACACCCCTTTAAGACATTGGTTCTAACA  
 a T L H S L H N L T R V W G N S V S Q D C -  
 b R Y T A F T T P E C G E I L A K I V -  
 c V T Q P S Q L N Q S V G K F C K P R L S -

2341 ----- 2400  
 CTGCGGAATATCGGTGGCATAAACATTTCACACTCCAAACAGCTCAAGAATGTCTCATCG  
 GACCGCTTATAGCGAACGTATTTGTAAAGTGTGACGTCTTCGACTTCTTACAGAGTACC  
 a L R N I R C I N I S H C N K L K N V S W -  
 b C G I S V A T F K T A T S R M S H G -  
 c A E Y P L H K H F T L Q Q A E E C L M G -

2401 ----- 2460  
 GTTCAGAACTCCCAAGCTAGAGGTGATTCAACTTTTCGACTCCAGAGAGATAGAGGAA  
 CAAGTCTTTGAGGCTTTCGATCTCTCACTAACTTCACAAGCTCACCTCTCTCTATCTCTT  
 a V Q K L P K L E V I E L F D C R E I E E -  
 b F R N S Q S R L N C S T A B R R N -  
 c S E T P K A R G D T V R L Q R D R G I -

2461 ----- 2520  
 TTGATAAGCGAACACGAGAGTCCATCCGTGGAAGATCCAACATTGTTCCCAAGCCTGAAG  
 AACTATTCGCTTGTCTCTCAGGTAGGCAGCTTCTAGCTTGTAAACAAGGCTTCGGACTTC  
 a L I S E H E S P S V E D P T L P P S L K -  
 b A N T R V H P S K I Q H C S Q A R -  
 c D K R T R E S I R R R S N I V F K P E D -

2521 ----- 2580  
 ACCTTCAGAACTAGCGATCTCCGAGAAGTAAACAGCATCTCTCCATCTCGATTTCATTC  
 TCGAAGCTTTGATCCCTAGACCGCTCTTCATTTCTCTAGGAGCGGTAGAGCTAAAGCTAAG  
 a T L R T R D L P E L N S I L P S R F S F -  
 b P E L G I C Q N T A S S H L D F H S -  
 c L E N G S A R T K Q H P P I S I F I P -

2581 ----- 2640  
 CAAAAGTTGAACATTAGTCAACAAATTCGCCAGAGTTAAGAACTGCCGTTTCAG  
 GTTTTCAACTTTCTAATCAGTAGTCTTTAACCGCGCTCTCAATTCTTTGACCGCAAGTC  
 a Q K V E T L V I T N C P R V K K L P P Q -  
 b K K L K H S S Q I A P E L R N C R F R -  
 c K S N I S H H K L P Q S E T A V S G -

2641 ----- 2700  
 GAGAGGAGGACCCAGATGAACCTTCCCAACAGTTTATTCTGAGGAGAAATGGTGGAAAGCA  
 CTCTCCTCTCGGCTTACTTGAACCGTTGTCAAAATAACACTCTCTTTTACCACCTTTCTCT  
 a E R R T Q M N L P T V Y C E E K W W K A -  
 b R G G P R T C Q Q F I V R R H G C K H -  
 c E E D P D E L A N S L L G E H V E S T -

FIG. 2 CONTINUED

```

2701 CTGAAAAGATCAACCAAGCAAGAGCTTGTATTACGGCGCTTTOTTCMAATTCA
-----
GACCTTTTCTAGTTCGTTTCCTCTCGAAACATAAATGGCGCGAAACAAGGTTAACT 2760
a L E R D Q P N E E L C Y L P R P V P N . -
b W K K I N Q T R S P V I Y R A L F Q I D -
c G K R S T K R R A L L F T A L C S K L I -

2761 TATAAGAGCTAAGAGCACTCTCTACAAATATGTTCATTCAAGTAGCAGGAAGCCAGGA
-----
ATATTCTCGATTCTCTGAGACATGTTTATACAGGTAAGTATTCATCTCTTCGGTCCT 2820
a Y K S . E H S V Q I C P F I S S R K P G -
b I R A K S T L Y K Y V H S . V A G S Q E -
c . E L R A L C T N H S I H X . Q E A R K -

2821 AGGTTGTTCCAGTGAAGTCATCAACTTTCCACATAGCCACAAAAGTACAGATTATGTAAT
-----
TCCAACAAGGTCAGTTCAGTAGTTGAAAGGTGTATCGGTGTTTTCATCTCTAATACATTA 2880
a R L F Q . S H Q L S T . F Q N . R L C N -
b G C S S E V I N F P H S H X T R D Y V I -
c V V P V K S S . F H : A T K L E I H . S -

2881 CATAAAACCAAGTATCCGCGA
-----
GTATTTTTCGTTTGATAGGCGCT 2903
a H K N Q T I R -
b I K T K L S A -
c . K P N Y P R -

```

Enzymes that do cut:

NONE

Enzymes that do not cut:

KpnI

FIG. 2 CONTINUED

-146  
 ATCGATTGATCTCTGGCTAGTCCGAGTAGTCTATTTCAGACCACTGCTAGCCCTCCCTCTC -146

CGSCATCATCGAGCTATTTCGAATTTTCGACGGTTATCGAATTCCTAGTCCGAACCCATT -146

CATCTCTCGAACCCACCACCGACCTTAAACAGCTCTCTGAGCCGATCATGAAAATT -146  
 MetLysIle

CGTCTAGTTGCTATAAATCAGACCTCTCTCAGCAGCGAGCTCTCTCAGACCGCCGACCT -146  
 AlaProValAlaIleAsnHisSerProLeuSerArgGluValProSerHisAlaAlaPro

ACTCAGCCAAAGCAACCAACTTCAATCTCAAGCTCCGCAATTCAGATCCAAAGAAAAGT -146  
 ThrGlnAlaLysGlnThrAsnLeuGlnSerGluAlaGlyAspLeuAspAlaArgLysSer

AGCCTCTAAGCCCGCAACCCGCTATTACTCTCTTAAAGACACTACTCCGACACAC -146  
 SerAlaSerSerProGluThrArgAlaLeuLeuAlaThrLysThrValLeuGlyArgHis

AAGATAGAGCTTCCGCTTTTCGAGGCTCTCTCAAAAGCAATCATCTAGCAGCAGAC -146  
 LysIleGluValProAlaPheGlyGlyTrpPheLysLysLysSerSerLysHisGluThr

CGCGTTCTAAGTCCCAACCCAGATATTTGAGCCTTCGCTCTCTCTTCTACCCAAAACCT -146  
 GlyGlySerSerAlaAsnAlaAspSerSerSerValAlaSerAspSerThrGluLysPro

TTCTTCGCTCTACCCACCTCTCTTACGTATCCAGCTTAACTAGCCAAATCGAATCTTCT -146  
 LeuPheArgLeuThrHisValProTyrValSerGlnGlyAsnGluArgMetGlyCysTrp

TATCCCTCCGCAAGCAATCTTCCGCAATCTCTCAAGCTCCGCTCTCTCTAGCCCTCTCT -146  
 TyrAlaCysAlaArgMetValGlyHisSerValGluAlaGlyProArgLeuGlyLeuPro

CAGCTTATCTAGCCAGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT -146  
 GluLeuTyrGluGlyArgGluAlaProAlaGlyLeuGlnAspPheSerAspValGluArg

TTTATTCAGATCAAGCAATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT -146  
 PheIleHisAsnGluGlyLeuThrArgValAspLeuProAspAsnGluArgPheThrHis

Figure 3.



10/25

GAAGACTTGGCTGCACTGTTGTATAAGCACGGCCGATTATATTGGCTGCAAACTCCC 615  
GUGLULeuGlyAlaLeuLeuTyrLysHisGlyProIleIlePheGlyTyrLysThrPro

AATGACAGCTGCCACATGTCGGTCTTCACTGGTGTGCAAAAGACAGCTGGCATTACT 635  
AsnAspSerTrpHisMetSerValLeuThrGlyValAspLysGluThrSerSerIleThr

TTTACGATCCCCCAGACGGGGCCGACCTAGCAATCCCGCTGATTACTTTAATAGCGA 755  
PheHisAspProArgGlnGlyProAspLeuAlaMetProLeuAspTyrPheAsnGlnArg

TTGGCATGGCAGGTTTCACACCCAAATGCTTACCCTAAGTACGAGGTAATTTACGTC 815  
LeuAlaTrpGlnValProHisAlaMetLeuTyrArgEnd

CGCCGATCATCACAAGCCCATCATGCCGCTACGCTACCTAAATGCTGCTGCTGCTGCT 875  
←→

CGTCTCTATTTCTGATCCGGAAGATGAGTCAAAATATCTGGCAAGAGCTGCTGCTGCT 935

CGACTCTTACGCTTCGGATGATCAGGTCCTTCCGACAGCGGCTTCTCATCAGCAT 995

CTGCCACAGCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055

CAACTGCCGCTTCCGATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115

AAAGACATAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1175

AAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235

CGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1295

CGTGGCAGGATTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1346

Figure 3 continued

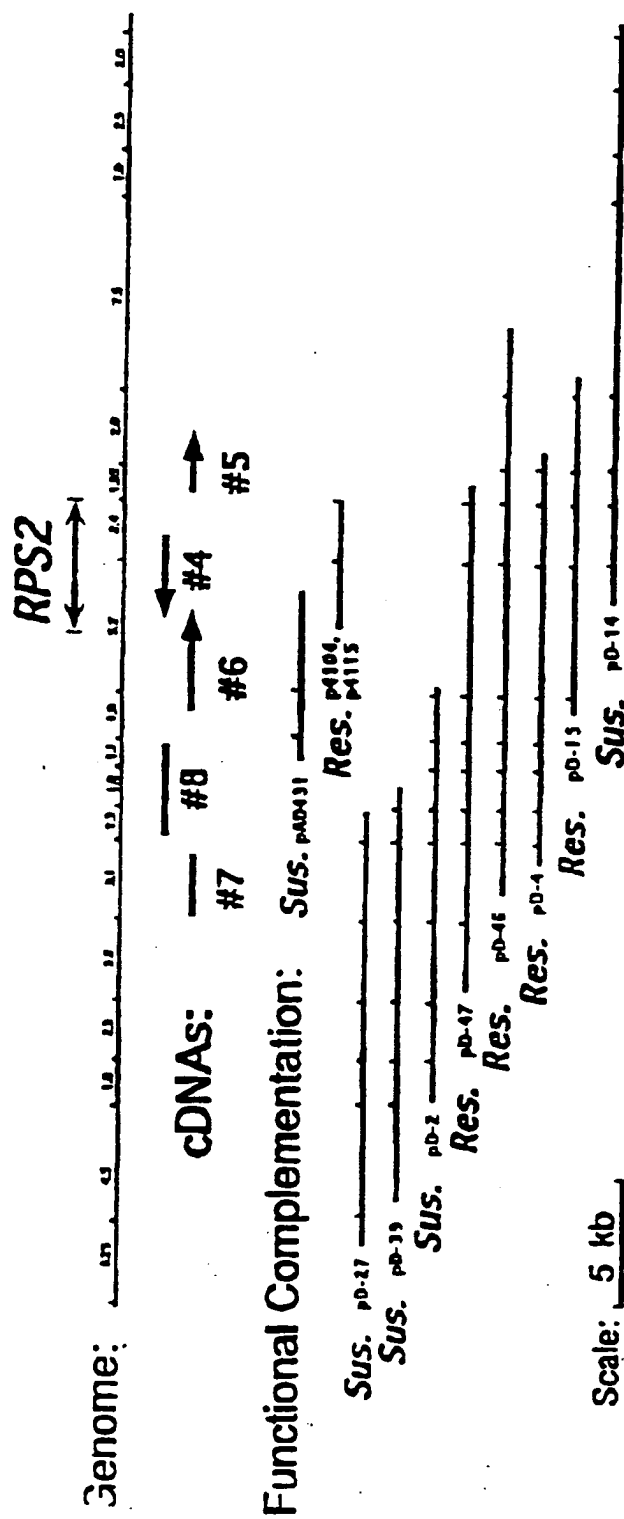


Figure 4.

12/25

FIGURE 5A

1 50  
 L6pro MSYLREVATA VALLPFIIL NOOWAFNSKO STVWSSOOS STVCAZROST  
 Nprot .....M  
 PrfP .....  
 rps2 .....

51 6 100  
 L6pro NPSGSIPEVE YEVTLSPROP DTREQFETL YQSLARYKIN TTRQODELLK  
 Nprot A8888888WS YEVTLSPROE DTATTTTSHL YEVLNDRKIK TTRQODKRLZY  
 PrfP ..... LRSKEDLIID LKHQIESVKE  
 rps2 ..... MDPIS8LIVG CAGVLCZM HAERROHKTO LQAITDLET

101 150  
 L6pro GKEIGPNLLR AIDQSKIYV IIS8GYADSK NCLMELAZIV RQZEDPRRI  
 Nprot CASIPGELCK AIEISQFATV VTSNYATBR NCLMELVKIM ECK.TAFKQ  
 PrfP OLLCLRPED NPSZYVEND ..... EA CGLIARVSVM AYKAE.....  
 rps2 AIGDLKARD OLTIRSQSG LICRS8CNRA REXLSAVQVT ERTA.....

151 7 200  
 L6pro IIPFIYMVDP SPVHQTCY KKAFRKHANK T..GGQTCN WQDALKVGO  
 Nprot VIFPIYDVP SHVTHQKESF AKAFTEENSTK YKQOVEGIOR WRIALNEAAN  
 PrfP .....VIDS CLAYSEPLNY KVLW..... ..IS..... EVLENIXLV  
 rps2 .....LLVR FRREORTAM RRY..... ..LSCTCCAD YELCKKVSAT

201 8 250  
 L6pro LKQNHIGQND KQQAIAQVS ADINSHISKE NLILE...TO EIVGTDDHET  
 Nprot LKESCDNRQK TDADCIHQIV DQISSKICKI SLSY....LO NIVGIDTHLE  
 PrfP KKVVGETGER RNTETVHEV AKITTVAPS FSAYTGRANE ZNEQFCSTED  
 rps2 LKSIQELREA SEATKIDGGS IQVICREIPI XSVVG..... ..NTRON

281 ① - P-loop 300  
 L6pro AVLEKLELDS ENVTWGLM GGGCKRTTA KAVYNKI... SSC.FDCC  
 Nprot KIEALLEIGT NGVRIMGIR GGGVQKTTA RAIFDTLLCA MOSSYQFDC  
 PrfP ELKDKLELGS PELDVTSIV GGGLCATTLA KRIYNDFEV... SRFQVHAQ  
 rps2 EQVLEPLSEE EERGIIIVG GGGVQKTTA Q8INNELLTX...S....MOY

Fig. 5A (cont)

301 350  
 L6pro CTIDNKRITQ ERDGVVVLQK KLVSEILRID ..SGSVOTPM CSGRRTIXE  
 Nprot CTIDNKRITQ ERDGVVVLQK KLVSEILRID ..SGSVOTPM CSGRRTIXE  
 P1P2 CWTCLYQWR EL..LITLND VLEP...S... ..DRMCKED GE..IADZLR  
 rna2 DVLIWVQMSR ET..GECTIQ AVGA...RLG ..LSWDEKET GENRALKIYR

351 400  
 L6pro KVARFYLIV LDDVEXKPF EDMIGSPKDF ISO..BRTYF SPMVVLGTL  
 Nprot KVARFYLIV LDDVEXKPF EDMIGSPKDF ISO..BRTYF SPMVVLGTL  
 P1P2 FLITRFLIL IDDVEXKPF ONLCHTFSD. VBRFLIIT TRLAEVAXV  
 rna2 ALROBELLIL IDDVEXKPF EXTVPRPD. RBRKRVMT TRLAEVAXV

401 450  
 L6pro NEN..OCHLYE VQMSKFRSL ELFSKRAFCK NT...PFSY YETLANDVVD  
 Nprot .EK..NDIYF VIALPDHESK OLFKOHAFCK EV...PNEN FEXLSITVW  
 P1P2 .KC..PBDPH LNLPRDESW TLLCKEVTGS E...SCPE LDDVGFZSK  
 rna2 .GA..EYK..LR VEFLEKHAM ELFSKRVWR OLEESSIR LAET...EVS

451 500  
 L6pro TIDGFLIKK VIGSLLPKOZ IAV..MEDTL ZOL....RT LHLDEVDRI  
 Nprot TIDGFLIKK VIGSLLPKOZ IAV..MEDTL ZOL....RT LHLDEVDRI  
 P1P2 SCRLFLSVV LVAQVLKQK KILDSKVVZ OSL..SOMI GSLEESISII  
 rna2 NCHFLILY TIDGAMAR R ETZEWIMAS EVLTPAEM KONTYVALL

501 550  
 L6pro KLSYDALPE AKELFLDIA CFFIGQ..NK ZEPYIMTDC NYYPASHIIF  
 Nprot KLSYDALPE AKELFLDIA CFFIGQ..NK ZEPYIMTDC NYYPASHIIF  
 P1P2 GFSYKNI..PH YLNCFLYFG GFLGKDIHD SMYKLVVAZ EYVQAM...  
 rna2 KLSYDNDSD LLRSCLYCA LTFEHSIET ZQLVETWGE GFLTSSHGVN

551 600  
 L6pro LIORCMIOVG .....DD DEFEMHDQLA DMREIVARE DVLFWKRSRI  
 Nprot LIORCMIOVG .....EY NGVGMHDLIQ DMREIVARE KD..PERSRI  
 P1P2 .....EK GEDTATYF ..LEASYW... ..  
 rna2 TTYKTYLIC DLRAACIIT ODEKTQVRM NVVSPALWM ASKQTYKEL

601 650  
 L6pro WSALEGIDLL LNRKGSXVK AISI..PHGVK YEFK..SECTL NLSILRYLHA  
 Nprot WSALEGIDLL LNRKGSXVK AISI..PHGVK YEFK..SECTL NLSILRYLHA  
 P1P2 ..... ..  
 rna2 ILVEPANGHT ZAPKADWRO ALVTSLEEN IQTL..PEKLI CPKLTTLIC

651 700  
 L6pro REAMLTODFN NLLPNLEWLE LFFYKNGEDD PPLNYTHON LII..VILNHS  
 Nprot REAMLTODFN NLLPNLEWLE LFFYKNGEDD PPLNYTHON LII..VILNHS  
 P1P2 ..... ..  
 rna2 QNSBLKRIPT OTFMPFVLR VLELSF... CSITTEPLSC NOL..VELYHL

## Fig 5A(cont)

```

L6pro HITADHWGCH RUDSGMAERL KVVRLASNYE LYGRVR... ..
Nprot .....YSL RULNTTETKHL PSL..... ...RID... ..
PrEP .....
rps2 SMSGTRISVL PQLLNLNL KHLDLORTCT LQTIFRDAIC WLSKLEVLNL

800
L6pro .LSD.CYRFP KSEVLSMGA IEMDEVIGE LKKLTLVLK PCPIQKINGE
Nprot .LSW.SKRLT RIFDFTGMEN LEY..VNLQ CSNLEEVNHS LGCCSKVIGL
PrEP .....
rps2 YYSY.AGWEL QSTGDEAKE LGFADLEYE NLTLIGITVL SLETLKTLF

850
L6pro TPCKLKLRE L.CSTFWGT NLREVVADIG QLSLKVLYK TCAGEVEINE
Nprot YLNDCKSLK P.....
PrEP .....
rps2 PQLLNLNL L.HVEICHEL LYFNLFLTN KGNLRLSL YSCMDEYLV

900
L6pro PFLOLK.... ..ELSTSR IFNLQLLDL ZVLKVTCCKD CFDPFPASP8
Nprot Y.LGLR.... ..SCDLKX LPFIYORCKV LI..... QIMMCGSGIR
PrEP .....
rps2 TPADPFWDL PSEVLTLRS LKHLTRVGN SVSQOCLRN RCINISHCNK

950
L6pro ZHSSVWVKV SKLKLGLKX TRHVVVDD ASSGHLPRY LLFTSLTYLK
Nprot ELPSIFQYK THVTLLE.. ..HSGOLVAL PSSICRL... ..KSLVLS
PrEP .....
rps2 LKGVSWQKL PKEVIELFD CRITZELIS MSFSEVDFP L.P.SLTLR

1000
L6pro IYQCTFTWL P.GIENLNL TSEVNDIPO TLGGDLGLG GLSLZLLRI
Nprot VSGCKLES L PEEICOLDNL RYDASDTL.. ..ILRP
PrEP .....
rps2 TDLPLNLBI LPSRFSQKY ETLVITNCPR VQULFQERR TUNLPTVYC

1050
L6pro RKVNGLARX GLKDLCSST CLKKTYITZ CPELZLLFC ELGVOTWVP
Nprot P.....SSI TRLNKLILM PRGFMGVNF EFPFVAZGLH
PrEP .....
rps2 ELKWKALKX GPFNEELCYL PRVFN....

1100
L6pro SMELTIRDC PRLEVGMIR SLKFFPMKX LELAVANIX EEDLDAIGSL
Nprot ELEYNL.SY CNLIDGOLPE EIGSLSLKX LELKONF.. EHLFSLAQ-
PrEP .....
rps2 .....

1150
L6pro ELVSLLEL DCTSSGIZRI VSSSKQYLT TLVWKVSLR EEDLLELKS

```

Fig. SA(cont)

Nprot	GAQSILDLK	.....	DCORLTCLE	ELPFELNELN	VECHMALKE
PrEP	.....	.....	.....	.....	.....
eps2	.....	.....	.....	.....	.....

	1151				1200
L6pro	LQDLYLEGCT	SLQRLEPLXL	KE.....LD	IGGCVDLTEL	VQTWAVPDL
Nprot	LDL.VTRRK	KLHRVKLODA	HNDTRYMLFA	YTFQNISSM	RHDISASDEL
PrEP	.....	.....	.....	.....	.....
eps2	.....	.....	.....	.....	.....

	1201				1250
L6pro	RQLTIRDCSR	LEVOPMIGSL	PKPFMLNELT	LSMVTITKED	ELIYLGSLER
Nprot	.SLIV.....	FTGQPYPERE	PSWPHHQWD	.SEVSVNLFE	MYIYDGLG
PrEP	.....	.....	.....	.....	.....
eps2	.....	.....	.....	.....	.....

	1251				1300
L6pro	LD.SLELTLD	STESSIERES	PLSKLCKETT	LIVEVPILRE	ESGLAEINEL
Nprot	FAVCYSRLI	DTAHLIPVC	.EDRMIRMTQ	KLALSECCTE	SENTSEWO..
PrEP	.....	.....	.....	.....	.....
eps2	.....	.....	.....	.....	.....

	1301				1350
L6pro	KLILYL.....	.....	.EGCTSLERL	WPDQOQLGSL	KNLNVLDIQG
Nprot	HFFPVFPAGL	WDTXANQNT	PMYGIITRLS	FSGLZKMTOL	RLLYKEGPEV
PrEP	.....	.....	.....	.....	.....
eps2	.....	.....	.....	.....	.....

	1351				1387
L6pro	CKSLBVDHLS	ALKTTLEPPA	RITWPDQPYR	.....	
Nprot	WALLCGRENS	NEPTENSTGI	RRTQYNNRTS	TYELINO	
PrEP	.....	.....	.....	.....	
eps2	.....	.....	.....	.....	

16 / 25

6

N 2 ASSSSSSRWSYVFLSFRGELTRKTFTSHLYEVLNDKGIXTFQDDKRLEY 51  
 L6 51 NPSGSFSPSVEYEVFLSFRGPDTRQFTDFLYQSLFRYKXHTFRDDDELLK 100  
 52 GATIPGELCKRAIEESOFATVVFSENATSRWCLNELVKIMECK.TRFKQT 100  
 101 GKEIGENLLRAIDQSKIYVPIISSGYADSKWCLMELAEIVRQEEPPRI 150  
 101 VPIFYDVDPSEVRNOKESFAKAFEEHETKYKDDVEGIQWRALNEAAN 150  
 151 ILPIFYMVDPSEVRHQTCCYKKAFAKXANKF..DGOTIQWWDALKKVG 198  
 151 LKGSCLNRDKTADCIQIVDQISSKLCKISLSY.LQNTVGIDTHLEKIE 199  
 199 LKGWHIGKNDKOGAIADKVSADTWSHISKENLILETDELVGIDHITAVL 248  
 200 SLEIGINGVRINGIWMGGVGKTTIARAIFDTLLGRMDSSYQFDGACFL 249  
 249 EKLSLDSSENVTMVGLYMGGIGKTTAKAVYNKI.....SSC.FDCCCFI 292  
 250 KDIKE..NKRGMHSLONALLSELLR...EKANYNNEEDGKHQMASRLRSK 294  
 293 DNIRETQEKGGVVLQKLVSEILRIDSGSVGFNNDSGGKRTIKERVSRF 342  
 295 KVLIVLDDIDKNDHYLEYLAGDLDFGNGSRITITTRDKHLI....EKND 340  
 343 KILVLDVDEKFKFEDMLGSPKDFISQ.SRFITTSRSMRVLGTINENQC 391  
 341 IIEYVETALPDHESIQLFKQHAFGKEVPNENFEKLSLEVWNYAKGLPLALK 390  
 392 KLYEVGSMKSPRSELEFSKHAFKNTPPSYETLANDVVDTTAGLPLTLK 441  
 391 VWGSLHLNRLTEWKSATHTMWN.SYSGIIDKLKISYDGLEPKQOEMFL 439  
 442 VIGSLFKOEIAVWEDTLEQLARTLNLEVDRLKISYDALNPEAKEIFL 491  
 440 DIACFLRGEEDYILQILESCHIGAEGYGLRILIDKSLVFISEYNQVOMHD 489  
 492 DIACFFIGONKEEPPYMWTDNCFYPASNIIFLIQRCMIQVGGDEDFKMD 541  
 490 LIQDMKYIVNFQKD.PGERSRLWLAKVEEVMSNNTGTMAEATWVSSY 538  
 542 QLRDMGREIVRREDVLPWKRSRISWAEEGIDLLNKGSSKVKALSI.PW 590  
 539 SSTLRFSSNOAVNMKRLRVFMGRSSSTHYAIDYLFNNLRCFVCTNYFW.. 526  
 591 GVKYEFKSECFNLSELRYLHAREAMLTGDFNNLLFNKWLLELPFYKHGE 640  
 587 .ESFPSTFELKMLVHLQLRH.....NSLRHLATETKHLPSL..... 621  
 641 DDPPLTNYTMQNLITVILESHITADDWGGWRHNMKAERLKVFLASNY 690  
 622 ....RRIDLSWSKRLTRIPDFTGMPNLEY..VNLYQCSNLEEVHSLGCC 665

FIGURE 5B

591 SLYGRVRLSDCWRFPKSTIEVLSMTATEMDEVOIGELKGLKTLVLKFCPI 740  
 666 SKVIGLYLNDCKSLKRFPCVNVESLEYLGLRSCDSLEKLPETIYGRMKP. 713  
 741 QKISGCTFGMLKGLREL.....CLEFNWGTNLRVWADIGQLSSLK 781  
 714 .....EIQIHMOGSGIRELP.SSIFQYKTHVTKLLDWM....KNLV 750  
 782 VLKTTGAKEVEINEFPLGLKELSTSSRIPNLSQLLDLEVLKVYDCKDGF 831  
 751 ALPSSICRLKSLVSLSVSCCKLESLEFEEIGDLNLRVFDASDTLILRP. 799  
 932 MPPASPSSESSVWVKV...SKLKSLEKTRINNVVDDASSGGHLPY 878  
 800 .....PSSIIRLNKLIILMFRGFKDGVHFEFPPVAE 830  
 879 LLPTSLTYLKYQCTEPTWLPGLNENLTSLEVNDIFQTLGGDLGLQ 927  
 831 GLHSLEYLNLSYCNLID..GGLPEET.GSLSSLKLDL..SRNFEHLPS 875  
 928 GLRSLEIIRIRKVNGLARIKGLKDLLCSSTCKLRFYITECPDLIELFC 977  
 876 STA....QIGALQSLDKDCQRLTQLPELPELNELHVDCHMAKFIHYL 921  
 978 ELGVOTVVVPSMAELTIRDCPRL.EVGFMIRSLPKFPM.....LKKDLA 1021  
 922 VTKRKLHRVKLDDAHNDTMYNLFAYTTFONISSMRHDISASDLSLTVF 971  
 1022 VANITKEEDLDAIGSLEELV..SLELELDDTSSGIERTVSSSKLOKLTL 1069  
 972 TGQFYPEKIPSWFHHQGDSSSVN.....LPENWYIFDKFLGFAVCY 1014  
 1070 VV.....KVPSLREIEGLEELKSLQDLYLEGCTSLGRLPLEKLKELDIGG 1114  
 1015 SRSLIDITTAHLIPVCDK.....MSRMTQKLA....LSECDTES 1049  
 1115 CPDLTELVTVVAVPSLRGLTIRDCPRLEVGFMIQSLPKFPMNLTLISM 1164  
 1050 SNYSEWDIHFFVFPFAGLWDTSKANGKTPNDYGIIRLSFSGEEIYGLRL 1099  
 1165 VNITKEDELEVLSLEELDSLELTLDOTCSSIERISF.LSKLOKLTTLIV 1213  
 1100 LYKEGPEVNALLQMPRENSNEPTEHSTGIRRTQYNNRTSFYELN 1143  
 1214 EVPSLREIEGLAELKSLRILYLEGCTSLER.LWPDQQQLGSLN 1256

FIGURE 5B (CON'T)



FIGURE 6

consensus	PXXaXX LXXLXXLXaXXXX aXXa	
505	PKAENW RQALVISLLD NR IQTL	
527	PEKLIK PK LTTLMQLQONSSLKKI	
550	PTGFFMHMPVLRVLDLSFTS ITEI	
574	PLSIKY LVELYHLSMSGTK ISVL	
597	PQELGN LRKLKHLDLQRTQFLQTI	
621	PRDAICWLSKLEVLNLYYSYAGWEL	QSFGEDEAEELG
658	FADLEY LENLTTLGITVLS LETL	KT
683	LFEFGALHKHIQHLHVEECHELLYF	NL
710	P SLTNHGRNLRRLSIKSDHLEYL	WT
736	PADFENDWLPSEVLTLHSLHNLTRY	WGN
765	SVSQDC LRNIRCINISHCNKLKNV	SWVQKL
795	PK LEV IELFDCREIEELISEHES	PSVED
823	PT LFPSLKTLRTRDLPENSI L	
845	PSRFS FQKVETLVITNCPRVKKL	

FIGURE 7

**FIGURE 8**

[illegible]

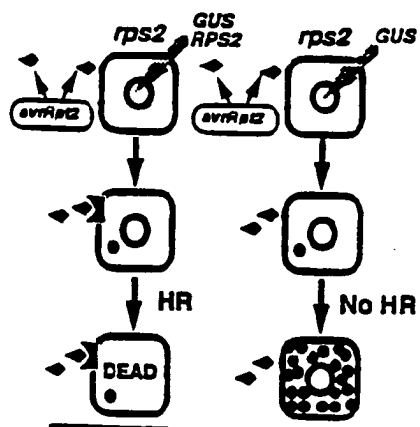
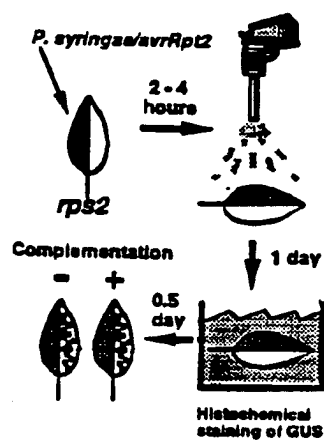
**RPS2 Transient Expression Assay****Principle of the assay****Actual procedure**

FIGURE 9

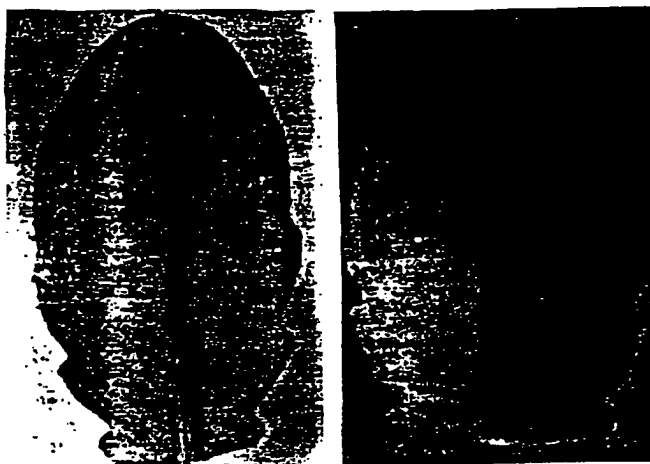


FIGURE 10

23 / 25

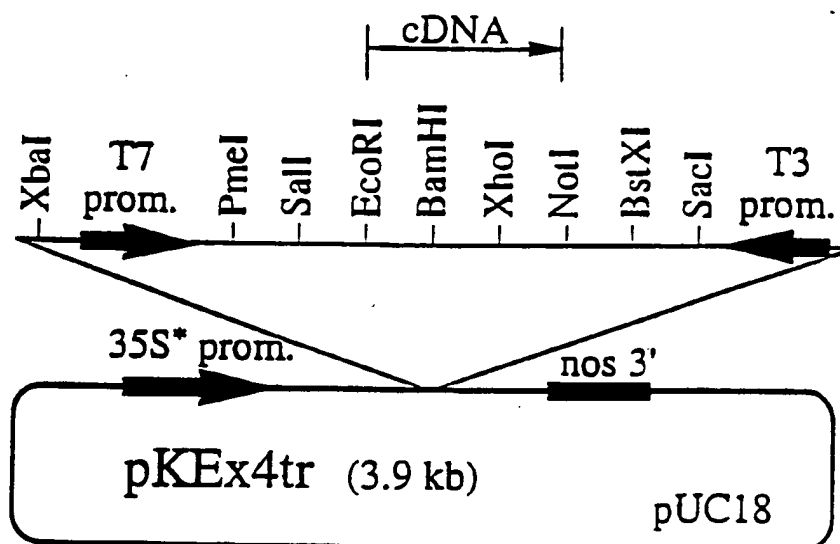


FIGURE 11

24 / 25

	10	30	40	50	60		
1	aaagcttaca	gattggatga	tctcttaacg	catgctgaag	tacttgcaca	aaagct-agca	60
61	atattcaagt	gttctcgtta	tgaattatttc	atgaacggaa	gcagcactga	gaaatcgagg	120
121	cccttggtat	ctgattttct	gcaagagatt	gagtcctgca	aggtagagtc	cagaaatgct	180
181	tgtttgcaag	tcttggtatc	atcacctttt	tccctgacag	atggagaaag	ccttgtaaat	240
241	ttcttattaa	aaaaccaggg	caaggtgccc	aatgatgatg	ctgttttttc	tqatggaaqt	300
301	ttagaggtat	caagcagcac	tqagaaatcg	ggacttccat	ctgattttct	ccgagagatt	360
361	gagtcctgtt	agataaaaga	ggccagaaaa	ttatatgata	aaatttttga	tgcacacat	420
421	tgttgaacga	gtaaagacga	tqgaaaaagc	tttatcaaca	ttatgttaac	ccaacagga	480
481	aaagtgctgg	actatgatgc	tggttcagtc	tcttatcttc	ttacccaaat	ctcagtagtt	540
541	aaagacaaac	tattgcacat	tggctcttta	cttgtagata	ttgtacagta	ccggaatatg	600
601	catatagaa	ttacagatct	cgtctgaacgt	gttcaagata	aaaactacat	tcgtttcttc	660
661	tctgtcaagg	gttatattcc	tgtcttggtat	tacacactat	atctctctga	tgtcaagcaa	720
721	tgtcttaagt	tgttgaggc	agaggtaaag	attatttctc	tgaaggtacc	agattcttca	780
781	agttatagct	tccctaaagc	aaatggatta	ggatattctc	attgtctttt	aggcaaatct	840
841	gaggtagctt	tacgttctaa	gctcgaattg	ataatcgact	taaaacatca	gattgaaatca	900
901	gtcaaggaag	gtctatttgc	cctaaagatca	ttcatttgatc	atttttcaag	aaagctatgt	960
961	gagcatgatg	aaagcttctg	tcttatagca	agagtttctg	taatggcata	caaggtctag	1020
1021	tatgtcattg	actcatgctt	ggcctattct	catccactct	ggtaacaaat	tctttggatt	1080
1081	tctgaagctc	ttgagaatct	taagcttctga	antaaagtgt	ttggggagac	atgtgaaaga	1140
1141	aggaacactg	aaagttactg	gcattgaagtt	gcaaaagacta	ccactaatgt	agcaccatct	1200
1201	ttttcagctt	atactcaaa	agcaaacgaa	gaaatggagg	gttttcaagga	tacaatagat	1260
1261	gaattaaaag	ataaactact	tggaggtatca	cctgagcttg	atgtcatctc	aatcgttggc	1320
1321	atgccaggtt	tgggcaagac	tacactagca	aaagaagattt	acaatgatcc	aaaggtcacc	1380
1381	tctcgtcttg	atgtccatgc	tcaatgtgtt	gtgactcaat	tatatctatg	gagagagttg	1440
1441	ttgttcacca	ttctgaatga	tgtgtcttgag	ccttctgact	gcaatgaaa	agaagatcga	1500
1501	gaaatagctg	atgagctacg	ccgtattttg	ttgaccaaga	gattcttgat	tcttactgat	1560
1561	gaggtgttgg	actataaagt	gtgggacaa	ttatgtatgt	gcttcaagta	tgtttcaaat	1620
1621	aggtatagaa	ttatcttaac	aaacctgttg	aatgatgtct	ccgaaatgt	caaatgtgaa	1680
1681	agtgtctccc	actcatcttg	tttattcaga	gatgacagaa	gttggaacat	attcagaaad	1740
1741	gaaatcttct	aaagagagag	ctgtccacct	gaacttgaag	atgtgggatt	tgaatatca	1800
1801	aaaggttgtt	gaggtgttgc	tctctcagtt	gtgttagttag	ctgtgttctt	gaaacagaaa	1860
1861	aaagagacac	tgaattcaig	gaaagttagta	gaacaaagtc	taagttccca	gaggtattgg	1920
1921	aggttggaa	agagcatatc	tataatttga	ttcagttacA	agaattttac	acactattct	1980
1981	aaagcttctt	ttctctattt	tGgaagattt	ttgcaaggaa	aggtatttca	tgaactcaaa	2040
2041	atgacaaagt	tgtgggttag	tgaagAgttt	gtacaagcaa	acaaacgaaa	aggaacagaa	2100
2101	gataccctga	caaggtttct	tggacgatct	tatttggtagg	aatctgttga	tggccatgga	2160
2161	gaagagacct	aaigccaaag	tgaanaagtg	ccgcattcat	gatttcttgc	ataaattctg	2220
2221	catggaaag	gccaaacaa	aggtatttct	tctccagatc	aataggttaa	aaaaactgta	2280
2281	ttaatcttac	attacaaaa	aaaagaactg	tatttaattt	actgtattat	gtttatgcca	2340
2341	actctcattt	ccatgtgttc	tctttcttct	aatccagtgg	agaaggtgta	tttccgtgac	2400
2401	gatttgaaga	ataccgattg	ttcgttcatt	cttaaccaag	tgaatctgat	ctgtggcgcc	2460
2461	catctcgttc	taagtctcgc	tctttactat	tcaatgcaat	tgtaccagat	aaactgttat	2520
2521	ggccgcgtga	tactctcttc	atttttgaga	gcttcaagct	tgttaagtg	tggactttgg	2580
2581	aatcattcaa	catgtgtgtt	acttttccca	ttgaaataca	atatctaat	cagatgaagt	2640
2641	actttggggc	ccaaactgat	gcaaatccaa	ttctttcacc	tatagctaa	cttgaaaaac	2700
2701	ttgagacttt	tgtcgttaag	ggatttggag	gagagatgat	attacattgt	tcaattctga	2760
2761	agatggtgaa	attgagggat	atacatgtaa	atgatcgggt	ttcttttgg	ttgcttgaga	2820
2821	acatggatgt	tttaactgg	aaactcaca	tacctaat	gaaactctt	tctactccgc	2880
2881	gtctctttta	tgttaagac	gcagagaaga	ttttgaggaa	gatgccaaaa	ttgagaaaa	2940
2941	tgaattgcat	attttcaggg	acatttgggt	attcaaggaa	actgaagggt	aggtgtgttc	3000
3001	gttttccca	actagatttt	ctaaagtcac	ttgagctcc	caagctgggt	tgaacagct	3060
3061	atccagccaa	acttctcacc	aaagttcaatt	tccctctgca	actaagggaa	ctgactttct	3120
3121	caaaagctcc	tctacattgg	acccaaattt	cgatcattgc	agaactgccc	aaactgttga	3180
3181	ttcttgaatt	attgttcaga	gcttttgaag	gggactcact	ggaagtgaaa	gattcagaat	3240
3241	tcttgaatct	caaatactta	aaacttgaca	acctcaaa	tgtacaatgg	cccatctctg	3300
3301	atgatgcttt	tcttaagctt	gaacatttgg	ttttaacgaa	atgttaagcat	cttgagaaaa	3360
3361	tcccttctcg	ttttgaagat	gctgtttgtc	tcaatagagt	tgaagtgaa	tgttgcaact	3420
3421	ggaatgttcg	caattcagcc	caagatattc	aaactatgca	acatgaagtt	atagcaaatg	3480
3481	atcaattcac	agttactata	cagcttccag	attgttctaa	agaacagccc	cttgactctt	3540
3541	agcaaaaggt	gttcttctgt	gtgttctatc	aaagtgcatt	aaactttatt	catcttctgt	3600
3601	tacaccaaga	catgtttatt	ttgtctagat	tacttgatac	attaaaaaa	atcgaactca	3660
3661	tatttctgct	acagttctaa	cttttcttg	gcttacttga	ggtctagatt	agatcaatgg	3720
3721	ttcatgtaat	tttaatttca	ctgttctatt	caactgtctt	atgatgttg	tgaatgaca	3780
3781	atattgttat	ccctagccaa	atttatttat	ttcaaatgaa	aaactgatgc	acaactactt	3840

FIGURE 12

25 / 25

```

3841 ttttctgaaa tgttttt  tttt .gcta  taaaattgac gaattgacag ctt  3900
3901 tttcagctaa actctttgac accagaaagt  tttttagaa  tactgtggtt ttatgaaaga 3960
3961 gttctgtaga attttatgct tttgcagaat  atagtcttaa acacacacac tttctgtttt 4020
4021 cagagatcagc agaaagttaa gttcaaggca  tttctgtttt ttttagaaca agtggagttc 4080
4081 ttatgttgaa tttttgaaaa gaagaaagaat  caggagcagg taaagtcttc tctttttatg 4140
4141 tttttttttt ttttagatgtt attctttctt  ctggaacgtg aacacccgtg aaagcatttt 4200
4201 aataaaaacc gagagaaaaa taagatcttt  ttatataaa  cattatcag taaatatgac 4260
4261 taattccta tggtagaact gtttgacaaa  atgatagaga ggggagtttt atagtataag 4320
4321 taanaacagg ttgagaaaaa aatccttgca  cyattttcaa tttctggcca catcacatg 4380
4381 tgtgtcaaa  tttctctctt taagtggaac  aagcaatcag aaagctcat tcttatcggc 4440
4441 gacataccaa taccagctga ctgtctcttc  ttggttaact tagccttgct tacttagact 4500
4501 attagattag ttactaatga actggttaat  tggaaaccaaa tgtagttagc ctgatgagct 4560
4561 ggtagacatg tatatatgaa gataacccgc  taactctagt cgtggttaa tttttcattt 4620
4621 tctatttttt tttttcacag agtatatatg  acctggccta aaagtttttg ctctactaat 4680
4681 ttaactatta ccgtggatga aacaaagcatg  gcaacatctt caacaactat cactcaagca 4740
4741 atgtaaaaaa tggaggttct acgagggta  catgtaaag  ttttctgcac acaagaggtt 4800
4801 ctgagacttg aacaaatcct gttcaaggca  gttgagatgc tagtaaaaga agaaagaat 4860
4861 gagcctgcac taattaatct cctgttatga  atgagagaat gagaaaaaga tggagcttca 4920
4921 tgaaccaaaa gttacctttt tttttcttc  ttaattggat tactttgaa  cacatgtttg 4980
4981 ttagtctgaa attgtaatgg tgaagtgttt  gtaaatatag gtagtgat  ttgaaagaat 5040
5041 ggttctgtta tttttacaaa ccggaatcat  ttcgtataa tttttctctg taatttttg 5100
5101 tttcggttta ttcattactc atttcagtaa  gctt  5136

```

FIGURE 12 (CON'T)



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/04589

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : C07K 14/415; C12N 15/29; C12Q 1/68

US CL : 536/23.1; 23.6, 24.3; 435/6; 530/370

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 23.6, 24.3; 435/6; 530/370

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MPSRCH, SDC, CAS ONLINE, IG

Search terms: SEQ ID NOS: 1-5, 105, 158, AND 191

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X ----- P, Y	Cell, Volume 78, issued 23 September 1994, S. Whitham et al., "The product of the tobacco mosaic virus resistance gene <i>N</i> : Similarity to Toll and the interleukin-1 receptor", pages 1101-1115, see sequences.	1, 10-13 ----- 14, 19-32
P, X ----- P, Y	Science, Volume 265, issued 23 September 1994, A.F. Bent et al., " <i>RPS2</i> of <i>Arabidopsis thaliana</i> : a leucine-rich repeat class of plant disease resistance genes", pages 1856-1860, see the entire document.	10-14, 19-32 ----- 1
P, X ----- P, Y	Cell, Volume 78, issued 23 September 1994, M. Mindrinos et al., "The <i>A. thaliana</i> disease resistance gene <i>RPS2</i> encodes a protein containing a nucleotide-binding site and leucine-rich repeats", pages 1089-1099, see the entire document.	10-14, 19-32 ----- 1

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reasons (as specified)	*Z*	document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means		
*P* documents published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search  
13 JULY 1995

Date of mailing of the international search report

04 AUG 1995

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

CHE SWYDEN CHERESKIN

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/04589

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Proceedings National Academy of Sciences, USA, Volume 90, issued December 1993, F. Bunz et al., "cDNAs encoding the large subunit of human replication factor C", pages 11014-11018, see peptide sequences.	14
X	Proceedings National Academy of Sciences, USA, Volume 90, issued December 1993, P.D. Burbelo, et al., "Cloning of the large subunit of activator 1 (replication factor C) reveals homology with bacterial DNA ligases", pages 11543-11547, see peptide sequences.	14
X	Biochem. Biophys. Res. Commun., Volume 193, Number 2, issued 15 June 1993, Y. Lu, et al., "Cloning and expression of a novel human DNA binding protein, PO-GA", pages 779-786, see peptide sequences.	14
Y	The Plant Cell, Volume 5, issued August 1993, B.N. Kunkel et al., "RPS2, an Arabidopsis disease resistance locus specifying recognition of <i>Pseudomonas syringae</i> strains expressing the avirulence gene <i>avrRpt2</i> ", pages 865-875, see the entire document.	1, 10-14, and 19-32
Y	Phil. Trans. R. Soc. Lond. B, Volume 342, issued 29 November 1993, C. Dean, "Advantages of Arabidopsis for cloning plant genes", pages 189-195, see especially Table 1.	1, 10-14, 19-32

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/04589

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1, 10-14, 19-32

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

**BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING**

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-32, drawn to RPS oligos and polypeptides.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:

1. The oligos of claim 1
  2. The oligos of claim 2.
  3. The oligos of claim 3.
  4. The oligos of claim 4.
  5. The oligos of claim 5.
  6. The oligos of claim 6.
  7. The oligos of claim 7.
  8. The oligos of claim 8.
  9. The oligos of claim 9.
- and
1. The peptides of claim 14.
  2. The peptides of claim 15.
  3. The peptides of claim 16.
  4. The peptides of claim 17.
  5. The peptides of claim 18.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: each group of oligos comprise a separate and distinct chemical entity which does not share a special technical feature with any other species. Likewise, each group of peptides comprise a separate and distinct chemical entity which does not share a special technical feature with any other species of peptide.

Group II, claims 33-37, drawn to identification of plant disease resistance using biolistics.

Group III, claim 38, drawn to an antibody.

Group IV, claim 39-40, drawn to the Prf amino acid sequence.

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I is drawn to specific products and the use of those products to isolate disease resistance genes while Group II is not drawn to the use of any specific product to isolate genes but rather to the use of biolistics to isolate genes. The use of biolistics is a different inventive concept than the use of the specific products of Group I. Therefore, Groups I and II do not share a special technical feature.

The inventions listed as Groups I and III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group III is drawn to an antibody which is a product that is chemically distinct from the products of Group I which are specific oligos and peptides. Therefore, Groups I and III do not share a special technical feature.

The inventions listed as Groups I and IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I is drawn to specific nucleotides and peptides related to RPS while Group IV relates to a Prf amino acid sequence. These appear to be separate and distinct chemical entities. Therefore, Groups I and IV do not share a special technical feature.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US95/04589

The inventions listed as Groups II and III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the antibodies of Group III could not be used in the method of Group II. Therefore, Groups II and III do not share a special technical feature.

The inventions listed as Groups II and IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the Prf amino acid sequence of Group IV could not be used in the method of Group II. Therefore, Groups II and IV do not share a special technical feature.

The inventions listed as Groups III and IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the antibody of Group III and the Prf amino acid sequence of Group IV appear to be separate, distinct and unrelated chemical entities. Therefore, Groups III and IV do not share a special technical feature.

Accordingly, the claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.